SRFP Internship Meeting Learnings

Page No.:

NEW MEETING

Date :- 24 May 2021

Q) :- N number of fishes are there in the pond we catch M fish and tag them. Second time $\frac{X}{L}$ caught fishes are tagged. find N

Sol^{n} A fair coin is tossed n times

X = No of heads obtained

$X \sim Bin(n, 1/2)$

$$f(x) = \binom{n}{x}\left(\frac{1}{2}\right)^n \quad , x \in \{0, 1, \dots n\}$$

~~similarly for~~
Catch-Recatch problem

M = No of tagged fish

N = Total No of fish

L = No of fish caught (2nd time)

X = No of tagged fish out of the L fish

$$f(x) = \frac{\binom{M}{x}\binom{N-M}{L-x}}{\binom{N}{L}} = \text{likelihood function}$$

N unknown

$= \mathcal{L}(N)$

We need a maximizer of $\mathcal{L}(N)$

$$\mathcal{L}(N+1) = \frac{\binom{M}{x}\binom{N+1-M}{L-x}}{\binom{N+1}{L}}$$

Page No.:

$$\frac{\ell(N+1)}{\ell(N)} = \frac{\binom{N-M+1}{L-x}\binom{N}{L}}{\binom{N-M}{L-x}\binom{N+1}{L}}$$

$$= \frac{(N-M+1)\,(N-L+1)}{(N-M-L+x+1)\,(N+1)}$$

$$= \frac{N^2 - NM + N - LN + LM - L + N - M + 1}{N^2 - NM - LN + Nx + N + N - M - L + x + 1}$$

$$\ell(N+1) \underset{<}{\geq} \ell(N)$$

iff $N^r \underset{<}{\geq} D^r$

after cancelling

$$LM \underset{<}{\geq} (N+1)x$$

$$\frac{LM}{x} \underset{<}{\geq} (N+1) \qquad , x > 0$$

$$\frac{LM}{x} - 1 \underset{<}{\geq} N$$

$$N \underset{>}{\leq} \underline{\underline{\frac{LM}{x} - 1}} \rightarrow \text{May be integer / May Not be integer}$$

case I :-

$\frac{LM}{x} - 1 = N_0 \rightarrow$ Integer

Suppose 27

Page No.:

if $N=0$: $L(1) > L(0)$ | $N=1$: $L(2) > L(1)$ | ... $N=26$: $L(27) > L(26)$

$N=27$: $L(27) = L(28)$ | $N=28$: $L(28) < L(27)$

$$L(0) < L(1) \ldots < L(26) < L(27) = L(28) > L(29)$$

$\Rightarrow$ $\because$ If $\frac{LM}{X} - 1$ is an integer N_0, then

$N = N_0, N_0+1$ are maximum estimates

Case 2:- Not Integer:

$\frac{LM}{X} - 1 = p \rightarrow$ not an integer

Suppose 27.6

if $N=0$: $L(1) > L(0)$ | $N=1$: $L(2) > L(1)$ | ... $N=26$: $L(27) > L(26)$

$N=27$: $L(28) > L(27)$ | $N=28$: $L(29) < L(28)$

$$L(0) < L(1) \ldots < L(26) < L(27) < \underline{L(28)} > L(29)$$

$\Rightarrow$ If $\frac{LM}{X} - 1$ is not an integer and is equal to p, then

$N = [p+1]$ is the maximum estimate where $[p+1]$ is the greatest integer function of p

—x—

Q) We have N numbered books in the library. We have same sample books $\{x_1, \dots x_n\}$ from which we have to estimate N.

Ans Books

$\{1, \dots \dots N\}$ → unknown

↓

sample of size n

$x_1, x_2, \dots x_n$

We want to get an estimate of N.

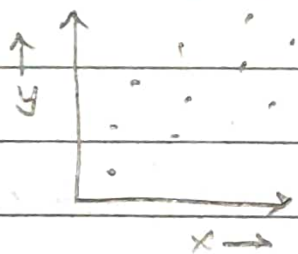
→ $\{X_1, X_2 \dots X_n\} \overset{iid}{\sim}$ Discrete Uniform (N)

$$f(x_1, x_2 \dots x_n) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{1}{N} \quad 1 \quad x_i \in \{1, \dots N\}$$

$$= \begin{cases} (1/N)^n & \text{if } x_i \in \{1, 2 \dots N\} \ \forall\, i \in \{1, \dots n\} \\ 0, & \text{otherwise} \end{cases}$$

Acc to MLE, $N = \max\{x_i\}$

Regression



Assumption

$y = f(x) + \varepsilon$

Regression :-

Aim at estimating $f(x)$

If we assum $f(x) = \alpha + \beta x$ → linear regression

given	predicted	actual
x_i →	$\alpha + \beta x_i$	y_i

error = residual = $(\alpha + \beta x_i - y_i) = r_i$

$$\Sigma (x_i)^2 = \Sigma (y - \alpha - \beta x_i)^2$$

↳ minimize (α, β)

Alternate

Consider $\Sigma |\_\_\_|$ not $\Sigma (\_\_)^2$

Now, if, $f(x) = \beta x$, we need to find β such that we minimize $\Sigma |\beta x_i - y_i|$

$$\Sigma |\beta x_i - y_i| = \underbrace{\sum_{x_i \neq 0} |\beta x_i - y_i|}_{\text{only considering this part as only this is dependent on } \beta} + \cancel{\sum_{x_i = 0} |\beta x_i - y_i|}^{\text{constant}}$$

Thus,

$$\sum_{x_i \neq 0} |x_i| \left| \frac{y_i}{x_i} - \beta \right| \rightarrow \text{minimize this}$$

↳ minimized when β is the weighted median of $\left(\frac{y_i}{x_i}\right)$ ie slope m_i

Q) How do we solve β to get minimum $\sum_{x_i \neq 0} |x_i| \left| \frac{y_i}{x_i} - \beta \right|$?

→ Suppose we have $\Sigma f_i |x_i - A|$ where f_i is frequency

x_i	f_i	$\sum_{j \leq i} f_j$ (cumulative frequency)
⋮	⋮	⋮

$\Sigma f_i |(x_i - A)|$ is minimized by x_i that lies on the median of the cumulative frequency. Thus 'A' is the weighted median of x_i

→ Minimizer of least square $\Sigma(y_i - \beta x_i)^2$

$\Sigma(y_i - \beta x_i)$ is minimized when,

$$\beta = \frac{\Sigma x_i y_i}{\Sigma (x_i)^2} = \frac{\Sigma (x_i)^2 \frac{y_i}{x_i}}{\Sigma (x_i)^2} = \frac{\Sigma (x_i)^2 m_i}{\Sigma (x_i)^2}$$

β is the weighted avg of the slope

→ Median is more robust against outlier as compared to mean

→ Minimization of absolute error is tougher than squared error.

Algo to find α, β to minimize absolute error

Minimize $\Sigma |y_i - \alpha - \beta x_i|$ wrt α & β?

→ Iterative method (Example)

initial → α_0

get β_0 by minimizing $\Sigma |y_i - \alpha_0 - \beta x_i|$

again α_1 by minimizing $\Sigma |y_i - \alpha - \beta_0 x_i|$

Repeat

OR

→ IRWLS → Iteratively reweighted least squares

→ Algo used for finding α, β to get minimum $\Sigma |\ \ |$

Task:- Develop an algo to find α, β ~~such that~~ such that $\Sigma |y_i - \alpha - \beta x_i|$ is minimum. Find it iteratively.

Page No.:

→ In regression we actually find $E(Y-\alpha-\beta X)^2$, but we dont know the distribution of y & x, hence we rely on the data point mean

we are minimizing $E(Y-\alpha-\beta x)^2$

$$= E_x\left(E_{Y|x}\left(Y-\underbrace{\alpha-\beta X}_{\text{let } f(x)=\alpha+\beta X}\right)^2\right)$$

⇒ Minimized when

$$f(x) = E(Y|X)$$

$$\alpha+\beta x_i = E(Y|X)$$

Similarly for $\Sigma|y_i-\alpha-\beta x_i|$ is minimized when

$$\alpha+\beta x_i = \text{Median}(Y|X)$$

$\Sigma\lvert y_i-mx_i-c\rvert$	Vs $\Sigma(y_i-mx_i-c)^2$
→ Difficult to find m & c to minimize the sum	→ Easy to find m & c to minimize the sum
→ $f(x)=mX+c$ is the Median $(Y\lvert X)$. This gives the minimum of the $\Sigma\lvert\_\rvert$	→ $f(x)=mX+c$ is the mean$(Y\lvert X)$. This gives the minimum of $\Sigma(\_)^2$
→ Not much affected by outliers	→ Gets easily affected by outliers

→ unstable solution	→ stable solution
→ possibility of multiple solutions	→ one solution

Task:- find ϕ such that $\Sigma \phi(x_i, A)$ is minimum when A is the p^{th} quantile of X

— x —

NEW MEETING

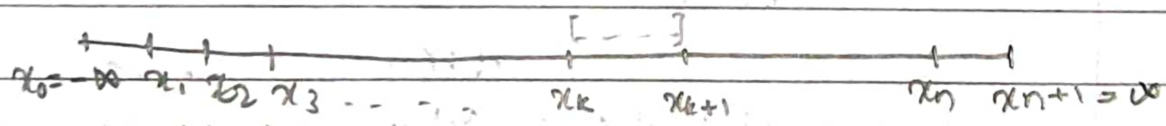
Date:- 2/6/2021

Q:- Proof for $\left[\underset{x}{\overset{\text{minimum}}{}} \Sigma \rho |(x_i - \theta)| \right]$ giving the p^{th} quantile $= \theta$

when $\rho = \begin{cases} 1-p, & x_i < \cancel{p^{th}\text{ quantile}}\ \theta \\ p, & x_i \geq \cancel{p^{th}\text{ quantile}}\ \theta \end{cases}$

$\Rightarrow$ $\psi(\theta)$ = function to be minimized $= \sum_{i=1}^{n} |x_i - \theta| + \sum_{i=1}^{n} (2p-1)(x_i - \theta)$

$= \sum_{i=1}^{n} \rho\, |x_i - \theta|$



Consider the interval $[x_k, x_{k+1}]$ & $x_k \leq \theta_1 < \theta_2 \leq x_{k+1}$

$$\psi(\theta_2) - \psi(\theta_1) = \sum_{i=1}^{k} (\theta_2 - x_i) + \sum_{i=k+1}^{n} (x_i - \theta_2) + (2p-1) \sum_{i=1}^{n} (x_i - \theta_2)$$

$$- \left\{ \sum_{i=1}^{k} (\theta_1 - x_i) + \sum_{i=k+1}^{n} (x_i - \theta_1) + (2p-1) \sum_{i=1}^{n} (x_i - \theta_1) \right\}$$

$$= \sum_{i=1}^{k} (\theta_2 - \theta_1) + \sum_{i=k+1}^{n} (\theta_1 - \theta_2) + (2p-1) \sum_{i=1}^{n} (\theta_1 - \theta_2)$$

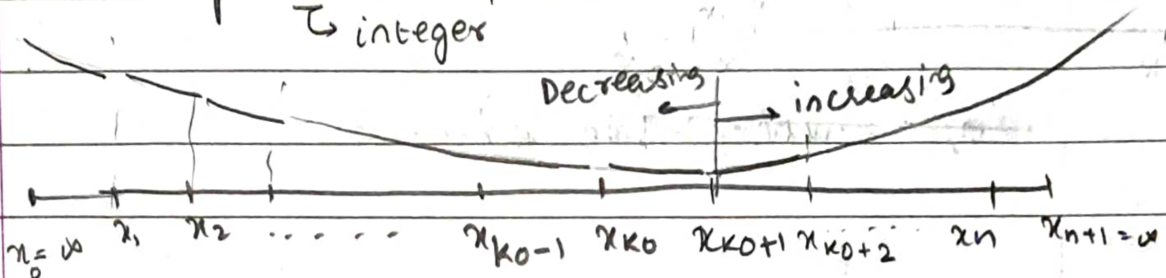
$$= k(\theta_2-\theta_1) - (n-k)(\theta_2-\theta_1) - (2p-1)n(\theta_2-\theta_1)$$
$$= (\theta_2-\theta_1)(k - \cancel{n} + k - 2pn + \cancel{n})$$
$$= (\theta_2-\theta_1)(2k-2np)$$
$$= 2\underbrace{(\theta_2-\theta_1)}_{+ve}\underbrace{(k-np)}_{\text{deciding factor}}$$

+ve

$$\boxed{\psi(\theta_2) \overset{>}{\underset{<}{=}} \psi(\theta_1) \text{ iff } k \overset{>}{\underset{<}{=}} np}$$

<u>case 1</u>:- np not an integer

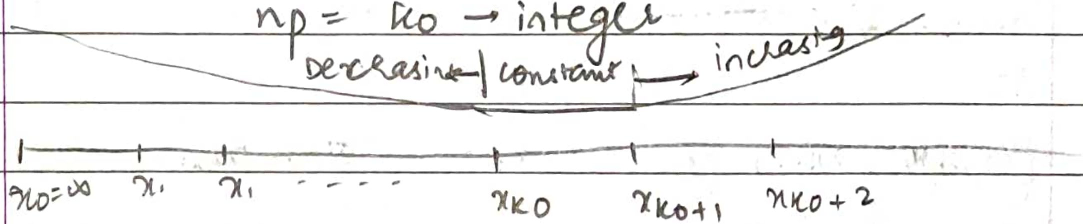
$np = \underset{\hookrightarrow \text{integer}}{k_0} + \delta$, $\delta \in (0,1)$



$\boxed{x_{k_0+1} \text{ is uniqe minimizer}} \rightarrow p^{th}$ quantile

<u>case2</u>:- np is an integer

$np = k_0 \rightarrow$ integer



$\boxed{\text{Any value in the interval } [x_{k_0}, x_{k_0+1}] \text{ gives the same global minimum}}$

~~mi~~

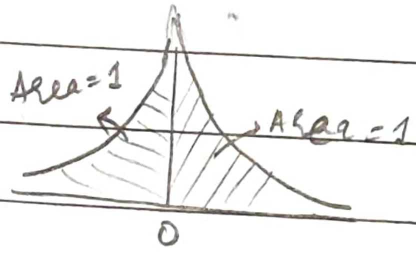
$\hookrightarrow p^{th}$ quantile

Q: Proof that LAD regression is MLE of its error.

$\Longrightarrow$ <u>Double exponential distribution</u>

$f(t) = \frac{1}{2} e^{-|t|}$

pdf



$\rightarrow \frac{1}{2}$ factor to make total Area=0

We have

$y_1 = \alpha + \beta x_1 + \varepsilon_1 \longrightarrow \varepsilon_1$ follows double exponential

$f(x_1, y_1) = f_x(x_1)\, f_{Y|X}(y_1)$

now

$y_1 - \alpha - \beta x_1 = \varepsilon_1$

$f(x_1, y_1) = f_x(x_1) \cdot \frac{1}{2} e^{-|\varepsilon_1|} = f_x(x_1) \cdot \frac{1}{2} e^{-|y_1 - \alpha - \beta x_1|}$

As

$(x_1, y_1), (x_2, y_2) \ldots\ldots (x_n, y_n)$ are iid's

$f(x_1, y_1, x_2, y_2 \ldots x_n, y_n) = \prod_{i=1}^{n} f(x_i, y_i)$

$L(\alpha, \beta) = \text{likelihood} = \prod_{i=1}^{n} f(x_i, y_i) = \prod_{i=1}^{n} f_x(x_i)\, f_{Y|X=x_i}(y_i)$

$= \prod_{i=1}^{n} f_x(x_i) \frac{1}{2} e^{-|y_i - \alpha - \beta x_i|}$

Distribution of x is independent of α & β, then MLE is obtained by maximizing

$$\prod_{i=1}^{n} e^{-|y_i-\alpha-\beta x_i|} = e^{-\Sigma|y_i-\alpha-\beta x_i|}$$

Maximized when $\sum_{i=1}^{n}|y_i-\alpha-\beta x_i|$ is minimized.

$\therefore$ Minimizing $\Sigma|y_i-\alpha-\beta x_i| \equiv$ MLE of joint dist when $\varepsilon \sim$ double exponential

$\boxed{\text{LAD Regression} \equiv \text{MLE when error is double exponentially distributed}}$

Now, if error has a distribution of $N(0,\sigma^2)$,

then we get

Minimizing $\Sigma(y_i-\alpha-\beta x_i)^2 \equiv$ MLE of joint dist when error $\sim N(0,\sigma^2)$

$\boxed{\text{least-square regression} \equiv \text{MLE when error is } N(0,\sigma^2) \text{ distributed}}$

LAD Regression

- 1. IRWLS
- 2. Stepwise Algorithm

1) **Step-wise Algorithm** (minimize $\sum |y_i - \alpha - \beta x_i|$)

Algorithm:

{ → initialize α_0, minimize $\sum |y_i - \alpha_0 - \beta x_i|$ wrt β to get $\beta = \beta_0$

→ put $\beta = \beta_0$, " $\sum |y_i - \alpha - \beta_0 x_i|$ wrt α to get $\alpha = \alpha_1$

Repeat until convergence }

Proof of convergence:-

Let $h(\alpha, \beta) = \sum |y_i - \alpha - \beta x_i|$

We know that

$h(\alpha_0, \beta) \geq h(\alpha_0, \beta_0) \geq h(\alpha_1, \beta_0) \geq h(\alpha_1, \beta_1) \geq \dots$

→ continuously decreasing, hence converges

2) **IRWLS for LAD Regression**

→ convergence is guaranteed

→ But cant be proved at this time

Task:- → Read MM algorithm & prove convergence of IRWLS for LAD regression as a spl case of MM algorithm

Page No.:

(Q1)

2 types of balls
↳ Red - - → p
↳ White - - → $1-p = q$
↳ proportions

urn

Draw a sample of size n

Note :- The urn contains large no. of of balls hence with / without doesnt matter.

But let us consider with replacement.

Draw a sample of size n
now

$n_1 \to$ Red , $n_2 = n - n_1 =$ White , MLE of p = ?

X = no. of red balls drawn $\sim Bin(n, p)$

$L(p) = \binom{n}{n_1} p^{n_1} q^{n-n_1}$ → difficult to directly minimize

$\log(L(p)) = \underbrace{\log\binom{n}{n_1}}_{\text{(constant)}} + n_1 \log p + (n-n_1)\log q$

$\frac{d}{dp}\log(L(p)) = 0 + \frac{n_1}{p} - \frac{(n-n_1)}{(1-p)}$

setting $\frac{d}{dp}\log(L(p)) = 0$

$\frac{n_1}{p} = \frac{(n - n_1)}{1-p} \Rightarrow n_1 - \cancel{n_1 p} = np - \cancel{n_1 p}$

$\Rightarrow \boxed{p = \frac{n_1}{n}}$

Page No.:

Naming done as

Q2)

R W	Pick two balls and put in a box	RR	p^2	n_1	no of boxes
		WW	q^2	n_2	
		RW	$2qp$	n_3	
				n	→ Total no of boxes

Now $p \to$ probability of picking a red ball

$q = 1-p \to$ probability of picking a white ball

We have many balls, hence we again consider a case with replacement.

$MLE(p) = ?$

Now

$$L(p) = \binom{n}{n_1, n_2, n_3} (p^2)^{n_1} (q^2)^{n_2} (2qp)^{n_3}$$

$$\log(L(p)) = \log\binom{n}{n_1, n_2, n_3} + 2n_1 \log p + 2n_2 \log q + n_3(\log 2 + \log p + \log q)$$

$$= \log\binom{n}{n_1, n_2, n_3} + (\log p)(2n_1 + n_3) + (\log q)(2n_2 + n_3) + n_3 \log 2$$

differentiate w.r.t 'p'

$$[\log(L(p))]' = 0 + \frac{(2n_1 + n_3)}{p} - \frac{(2n_2 + n_3)}{q} + 0$$

setting $[\log(L(p))]' = 0$

$$\frac{2n_1 + n_3}{p} = \frac{2n_2 + n_3}{1-p} \Rightarrow p = \frac{2n_1 + n_2}{2n} = \frac{\text{No of red balls}}{\text{Total balls picked}}$$

naming done as

Q5) RW OO $\xrightarrow{\text{Put two balls and put in a box}}$ RR, WR $\to$ with red — n_1; WW $\to$ without red — n_2 } no of boxes

$n \to$ Total no. of boxes

$p \to$ probability of picking a red ball
$q = 1-p \to$ " " " " white "

We have many balls, hence we consider the case with replacement

$MLE(p) = ?$

now

$$\text{Likelihood of } q^2 = \binom{n}{n_2} (q^2)^{n_2} (1-q^2)^{n_1}$$

↳ similar to binomial

∴ we know that

$$MLE \text{ of } q^2 = \frac{n_2}{n}$$

now, if $\hat{\theta}$ is MLE of θ, then $\tau(\hat{\theta})$ is MLE of $\tau(\theta)$ where τ is a function of θ

$$\therefore MLE \text{ of } \sqrt{q^2} = \sqrt{\frac{n_2}{n}} = MLE \text{ of } q$$

$$\therefore MLE \text{ of } 1-q = \boxed{MLE \text{ of } p = 1 - \sqrt{\frac{n_2}{n}}}$$

Page No.:

Q.4) ABO Blood Group System

→ probability of Genotype A — p] codominant
" " " B — q
" " " O — r] recessive

$p+q+r=1$

Genotype	Phenotype	Probability	no of people
AA, AO	A	p^2+2pr	n_1
BB, BO	B	q^2+2qr	n_2
OO	O	r^2	n_4
AB	AB	$2pq$	n_3

n → Total people in sample
↓
known

p, q, r → unknown (have to estimate)

now

$$L(p,q,r) = \binom{n}{n_1, n_2, n_3, n_4}(p^2+2pr)^{n_1}(q^2+2qr)^{n_2}(r^2)^{n_4}(2pq)^{n_3}$$

$$\log(L(p,q,r)) = \log c + n_1\log(p^2+2pr) + n_2\log(q^2+2qr) + n_4\log(r^2) + n_3\log(2pq)$$

How to estimate p, q, r? We can set

$$\frac{\partial}{\partial q}\log(L(p,q,r)) = \frac{\partial}{\partial p}\log(L(p,q,r)) = 0$$

[$r = 1-p-q$, hence no need for $\partial/\partial r$]

Page No.:

$$\frac{\partial}{\partial q}(\log(L(p,q,r))) = \frac{\partial}{\partial p}(\log(L(p,q,r))) = 0$$

Task ↳ Use Newton - Raphson Method (or) → Code it out
↳ EM Algorithm

Hint :-

Set $p = 0.3$, $q = 0.45$, $r = 0.25$

		Cumulative bounds
A	p^2+2pr	a_1
B	q^2+2qr	a_2
AB	$2pq$	a_3
O	r^2	a_4

→ Generate random number b/w (0,1)
→ According to cumulative bounds set the blood group as A, AB, B or O
→ Repeat this for 1000 times

Now forget p, q & r. Estimate them through EM or NR method to obtain $\hat{p}, \hat{q}, \hat{r}$

Compare with initial chosen values of p, q & r.

For NR method

$$\frac{\partial}{\partial q}\log(L(p,q,r)) = \frac{n_2}{q^2+2qr}(2q+2r) + \frac{n_3}{q} = 0$$

$$\frac{\partial}{\partial p}\log(L(p,q,r)) = \frac{n_1}{p^2+2pr}(2p+2r) + \frac{n_3}{p} = 0$$

$$\frac{\partial}{\partial r}\log(L(p,q,r)) = \frac{n_1}{p^2+2pr}(2p) + \frac{n_2}{q^2+2qr}(2q) + \frac{2n_4}{r} = 0$$

$\Delta n_1 = 2p' + 2r$ | $\Delta n_2 = 4q + 2r$ | $\Delta n_4 = 2r$

$$\frac{\partial}{\partial q} = \frac{2n_2}{q^2+2q(1-q-p)}(\cancel{q}+1-\cancel{q}-p) + \frac{n_3}{\cancel{q}} =$$

$$\frac{2n_2(1-p)}{\cancel{q}\,2-2p-q} + n_3 = 0 \quad \cancel{\times\times\times} / \cancel{q}$$

$$2n_2(1-p) + n_3(2-2p-q) = 0 \longrightarrow f$$

$$\frac{\partial}{\partial p} = \frac{2n_1(1-q)}{2-2q-p} + n_3 = 0 \quad \cancel{\times\times} \;/ q$$

$$2n_1(1-q) + n_3(2-2q-p) = 0 \longrightarrow g$$

$$J = \begin{bmatrix} \frac{\partial f}{\partial p} & \frac{\partial f}{\partial q} \\ \frac{\partial g}{\partial p} & \frac{\partial g}{\partial q} \end{bmatrix} = \begin{bmatrix} -2n_2 - 2n_3 & -n_3 \\ -n_3 & -2n_1 - 2n_3 \end{bmatrix}$$

$$\begin{bmatrix} p \\ q \end{bmatrix}^{i+1} = \begin{bmatrix} p \\ q \end{bmatrix}^{i} - \underset{2\times 2}{\begin{bmatrix} J \end{bmatrix}^{-1}} \underset{2\times 1}{\begin{bmatrix} f^i \\ g^i \end{bmatrix}}$$

This is giving r $-ve$. → clearly this didnt work

$$\frac{\partial}{\partial q} = \frac{2n_2(q+r)}{\cancel{q}(q+2r)} + \frac{n_3}{\cancel{q}} = 0$$

$$2n_2(q+r) + n_3(q+2r) = 0 \longrightarrow f$$

$$\frac{\partial}{\partial p} = 2n_1(p+r) + n_3(p+2r) = 0 \longrightarrow g$$

$$h = \quad p+r+q-1 = 0 \longrightarrow h$$

$$J = \begin{bmatrix} \partial f/\partial p & \partial f/\partial q & \partial f/\partial r \\ \partial g/\partial p & \partial g/\partial q & \partial g/\partial r \\ \partial h/\partial p & \partial h/\partial q & \partial h/\partial r \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 2n_2+n_3 & 2n_2+2n_3 \\ 2n_1+n_3 & 0 & 2n_1+2n_3 \\ 1 & 1 & 1 \end{bmatrix}$$

$n_1 = nA$

$n_2 = nB$

$n_3 = nAB$

$$\begin{bmatrix} p \\ q \\ r \end{bmatrix}^{i+1} = \begin{bmatrix} p \\ q \\ r \end{bmatrix}^{i} - \begin{bmatrix} J \end{bmatrix}^{-1}_{3\times 3} \begin{bmatrix} f^i \\ g^i \\ h^i \end{bmatrix}_{3\times 1}$$

This also results in -ve r Here also NR didnt work.

→ Later I found that $\frac{\partial(\log L)}{\partial q}$, $\frac{\partial(\log L)}{\partial r}$ & $\frac{\partial(\log L)}{\partial p}$ are not going to '0' at the required p, q & r. Hence minimizing $\frac{\partial}{\partial q/\partial p/\partial r}$ is not going to give the maximum of log(L) as there is no global minima.

NEW MEETING | 9^{th} June 2021 | 6 pm

Page No.:

Q) Consider the same Q as (Q3) in the previous meeting. E-M algorithm approach?

Ans →

Boxes Assume

p^2	RR	n_1	n_{11}
$2pq$	RW		n_{12}
q^2	WW	n_2	

n ↳ known; n_{11}, n_{12} ↓ unknown

where $n_{11} + n_{12} = 1$

MLE of $p = ?$

$$L(c) = \binom{n}{n_{11}, n_{12}, n_2} (p^2)^{n_{11}} (2pq)^{n_{12}} (q^2)^{n_2}$$

|| n_{11} & n_{12} taken to separate p^2 + $2pq$

↳ complete likelihood function

$$\hat{p} = \frac{2n_{11} + n_{12}}{2n} \longrightarrow \text{MLE of } p$$

But, n_{11} & n_{12} → unknown

$$\log(L(c)) = \log c_0 + 2n_{11}\log p + 2n_2 \log q + n_{12}\log 2 + n_{12}\log p + n_{12}\log q$$

Iterative Algorithm:-

→ Start with $p = p_0$, $q = 1 - p_0$ (Better to start with a rough estimate of p_0)

E-step → Take $E(\log L(c) |$ given n_1 & $n_2)$

$$= E[(\log c_0 + n_{12}\log 2 + 2n_{11}\log p + n_{12}\log p + 2n_2 \log q + n_{12}\log q) | n_1, n_2]$$

$$= C_1 + \log p(2E(n_{11}|n_1, n_2) + E(n_{12}|n_1, n_2)) + \log q(2n_2 + E(n_{12}|n_1, n_2)) +$$

Here ($C_1 = \log c_0 + n_{12}\log 2$)

Page No.:

Find

$$E_{p^0,q^0}(n_{11} \mid n_1, n_2) = n_{11}^0$$
$$E_{p^0,q^0}(n_{12} \mid n_1, n_2) = n_{12}^0 = n_1 - n_{12}^0$$
(here)

Now

$n_{11}^0 = ?$, we see that $n_{11} \mid n_1, n_2 \sim \text{Bin}\left(n_1, \frac{p^2}{p^2+2pq}\right)$

$$\therefore E_{p^0,q^0}(n_{11} \mid n_1, n_2) = n_{11}^0 = (n_1)\left(\frac{(p^0)^2}{(p^0)^2 + 2p^0q^0}\right)$$

Similarly

$$n_{12}^0 = n_1 \frac{(2p^0q^0)}{(p^0)^2 + 2p^0q^0}$$

Now

$$E_{p^0,q^0}(\log L(c) \mid n_1, n_2) = C_1 + \log p\,(2n_{11}^0 + n_{12}^0) + \log q\,(2n_2 + n_{12}^0)$$

M-Step → Maximize $E_{p^0,q^0}(\log L(c) \mid n_1, n_2)$ wrt p to get $p^{(1)}$

We know $p^1 = \frac{2n_{11}^0 + n_{12}^0}{2n}$, $q^1 = \frac{n_{12}^0 + 2n_2}{2n}$ $\left(\text{using } \frac{d}{dp}(E) = 0\right)$

Iteration → Repeat till convergence, that is when

$$|(p^{r+1} - p^r)| < \varepsilon \quad \text{or} \quad |q^{r+1} - q^r| < \varepsilon$$

where $\longrightarrow 0 < \varepsilon << 1$
predefined

Generelized M-step (for this problem)

$$p^{r+1} = \frac{2n_{11}^{(r)} + n_{12}^{(r)}}{2n}, \quad q^{r+1} = \frac{n_{12}^{(r)} + 2n_2}{2n},$$

$$n_{11}^{(r)} = n_1 \frac{p^{(r)2}}{(p^{r})^2 + 2p^{(r)}q^{(r)}}, \quad n_{12}^{(r)} = n_1 \frac{2p^{(r)}q^{(r)}}{p^{(r)2} + 2p^{(r)}q^{(r)}}$$

or

$$n_{12}^{(r)} = n_1 - n_{11}^{(r)}$$

Convergence of this Algorithm

$$q^{(r+1)} = \frac{1}{2n}\left(\frac{2q^{(r)}\cancel{p^{(r)}}\, n_1}{p^{(r)2} + 2\cancel{p^{(r)}}q^{(r)}} + 2n_2\right)$$

$$= \frac{1}{2n}\left(\frac{2q^{(r)} n_1}{p^{(r)} + 2q^{(r)}} + 2n_2\right) = \frac{n_1}{n}\frac{q^{(r)}}{1+q^{(r)}} + \frac{n_2}{n}$$

We have

$q = \phi(q)$ where $\phi(q) = \frac{n_1}{n}\frac{q}{1+q} + \frac{n_2}{n}$

$$\phi'(q) = \frac{n_1}{n}\left(\frac{1}{1+q}\right)^2 \rightarrow |\phi'(q)| < 1 \quad \forall q$$

when an iterative algorithm follows a scheme of $y = \psi(y)$ and $|\psi'(y)| < 1$ it always converges//

Page No:

Convergence of $|\phi'(x)| < 1$

Suppose we want to solve $x^2 - x - e^x = 0$

$$x = \sqrt{x + e^x} = \phi(x)$$

Note :- ϕ is not unique

we choose ϕ such that $|\phi'(x)| < 1$

Start with x_0, $\Rightarrow x_1 = \phi(x_0) \Rightarrow x_2 = \phi(x_1) \ldots$

Now, as $|\phi'(x)| < 1$, convergence is assured

Lets prove this

Let x^* be the true soln of $x = \phi(x)$ i.e $x^* = \phi(x^*)$

$$x_n - x^* = \phi(x_{n-1}) - \phi(x^*)$$
$$= (x_{n-1} - x^*)\,\phi'(\xi_1) \qquad \xi_1 \in (x_{n-1}, x^*)$$

$$|x_n - x^*| = |x_{n-1} - x^*|\,|\phi'(\xi_1)|$$
$$\leq |x_{n-1} - x^*|\,M \leq |x_{n-2} - x^*|\,M^2 \ldots\ldots$$
$$\ldots \leq |x_0 - x^*|\,M^n$$

where $\max_x |\phi'(x)| = M < 1$

$\because |\phi'(x)| < 1 \ \forall x$

now,

as $n \to \infty$

$M^n \to 0$ as $M<1$ $\Rightarrow$ $|x_n - x^*| \to 0$

$\therefore$ Convergence is assured when $|\phi(x)| < 1$ for all x

---

Does E-M converge to the right value?

we know E-M converges $\rightarrow$ but is it a valid estimate?

Now,

$$q = \phi(q) \Rightarrow q = \frac{n_1}{n}\frac{q}{1+q} + \frac{n_2}{n}$$

$$\Rightarrow nq(1+q) = n_1 q + n_2(1+q)$$

$$\Rightarrow \cancel{nq} + nq^2 = \cancel{n_1 q} + n_2 + \cancel{n_2 q}$$

$$\Rightarrow q^2 = \frac{n_2}{n}$$

$$\Rightarrow q = \sqrt{\frac{n_2}{n}}$$ $\longrightarrow$ This is what is the true solution

Hence EM converges to the valid solution.

— x —

Page No.:

Q) Consider the ABO blood group problem in the previous meeting (Q4). Solving it using E-M algorithm.

→

AA AO	Blood Group	Samples Known	unkown
$p^2 + 2pr$	A	n_1	n_{11} (AA Gen), n_{12} (AO gen)
BB BO $q^2 + 2qr$	B	n_2	n_{21} (BB Gen), n_{22} (BO gen)
AB $2qp$	AB	n_3	
OO r^2	O	n_4	
		n (Total)	

$$L(c) = \binom{n}{n_{11}, n_{12}, n_{21}, n_{22}, n_3, n_4} (p^2)^{n_{11}} (2pr)^{n_{12}} (q^2)^{n_{21}} (2qr)^{n_{22}} (2pq)^{n_3} (q^2)^{n_4}$$

$$\log L(c) = c_0 + (2n_{11} + n_{12} + n_3)\log p + (2n_{21} + n_{22} + n_3)\log q + (n_{12} + n_{22} + 2n_4)\log r$$

Algorithm

→ Start with $p^0, q^0, r^0 = 1 - p^0 - q^0$

E-step → $E(\log L(c) \mid n_1, n_2, n_3, n_4) = c_0 + \log p\,(2E(n_{11} \mid n_1, \dots, n_4) + E(n_{12} \mid n_1, \dots, n_4) + n_3)$
$+ \log q\,(2E(n_{21} \mid n_1 \dots n_4) + E(n_{22} \mid n_1, \dots n_4) + n_3)$
$+ \log r\,(E(n_{12} \mid n_1 \dots n_4) + E(n_{22} \mid n_1 \dots n_4) + 2n_4)$

Page No.:

$E_{p^0,q^0,r^0}(\log L(c) \mid n_1, \dots n_4)$

$$= C_0 + (2n_{11}^0 + n_{12}^0 + n_3)\log p + (2n_{21}^0 + n_{22}^0 + n_3)\log q$$
$$+ (n_{12}^0 + n_{22}^0 + 2n_4)\log r$$

• Expectation of $n_{11}, n_{12}, n_{21}, n_{22}$ given $p^0, q^0, r^0, n_1, n_2, n_3$ & n_4

$$n_{11}^0 = n_1 \frac{(p^0)^2}{(p^0)^2 + 2p^0r^0}, \quad n_{12}^0 = n_1 - n_{11}^0$$

$\left(\text{as } n_{11} \sim \text{Bin}\left(n_1, \frac{(p)^2}{(p)^2 + 2pr}\right)\right)$

Similarly

$$n_{21}^0 = n_2 \frac{(q^0)^2}{(q^0)^2 + 2q^0r^0}, \quad n_{22}^0 = n_2 - n_{21}^0$$

M step $\longrightarrow$ Maximize $E_{p^0,q^0,r^0}(\log L(c) \mid n_1, n_2, n_3, n_4)$

$$p^{(1)} = \frac{2n_{11}^{(0)} + n_{12}^{(0)} + n_3}{2n} \longrightarrow \frac{\text{Total A alleles}}{\text{Total (A, B, O) alleles}}$$

(AA considered as 2A)

$$q^{(1)} = \frac{2n_{21}^{(0)} + n_{22}^{(0)} + n_3}{2n}$$

$$r^{(1)} = \frac{n_{12}^{(0)} + n_{22}^{(0)} + 2n_4}{2n}$$

Iteration $\longrightarrow$ Repeat E & M step till convergence

we can use $\left(|p^{n+1} - p^n|, |q^{n+1} - q^n|, |r^{n+1} - r^n|\right) < \varepsilon$

Page No.:

we sometimes also use

$$\left(\left| \frac{p^{n+1} - p^n}{p^n} \right| , \left| \frac{q^{n+1} - q^n}{q^n} \right| , \left| \frac{r^{n+1} - r^n}{r^n} \right| \right) < \varepsilon$$

where $0 < \varepsilon << 1$ → predefined value

Here it is not straight forward to show the convergence of EM algo to the desired value.
we will see that later (when we develop a general theory)
But, through intuition from previous problem, we have developed this general theory.

Task :-

Code out the solution of ABO problem using
→ EM Algorithm
→ NR "
→ Fisher scoring method which is a modified version of N-R method.

we sometimes also use

$$\left(\left| \frac{p^{n+1} - p^{n}}{p^{n}} \right|, \left| \frac{q^{n+1} - q^{n}}{q^{n}} \right|, \left| \frac{r^{n+1} - r^{n}}{r^{n}} \right| \right) < \varepsilon$$

where $0 < \varepsilon << 1$ → predefined value

Here it is not straight forward to show the convergence of EM algo. to the desired value
We will see that later (when we develop a general theory)
But, through intuition from previous problem, we have developed this general theory.

Task :-

Code out the solution of ABO problem using
→ EM Algorithm
→ NR "
→ Fisher ~~Bo~~ scoring method which is a modified version of N-R method.

Q. We have two normal Distributions → $N(0,1)$
→ $N(1,1)$

We have a coin with $P(H) = p$ & $P(T) = 1-p = q$

If we get a H → we pick a point from $N(0,1)$

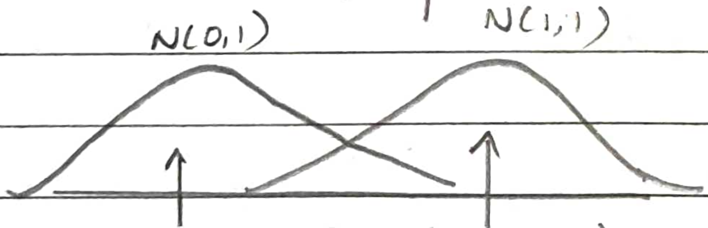
" " " " T → we pick a point from $N(1,1)$

We have n such points.

Now, we know → $x_1, x_2, \ldots x_n$ (points)
→ $N(0,1)$, $N(1,1)$ (distribution)

What is MLE of p? (p & q are unknown)

<u>Solution</u> →



$p = P(H)$ $q = P(T)$ → coin tossed

<u>points</u> :- $x_1, x_2, \ldots \ldots x_n$

Method of Moments estimate ⇒ $\bar{x} = \frac{0 \times p + 1 \times q}{p+q} = q$

Let $\boxed{q = \bar{x}}$ be the initial guess obtained through MOM estimate.

<u>Original Data</u> $x_1, x_2 \dots x_n$

<u>Complete Data</u> $\binom{x_1}{y_1}, \binom{x_2}{y_2} \dots \binom{x_n}{y_n}$

where $y_i = \begin{cases} 0 & \text{if } x_i \in N(0,1) \\ 1 & \text{"} \; x_i \in N(1,1) \end{cases}$

But $y_1, y_2 \dots y_n$ are unknown, hence we use E-M Algorithm

$$L(c) = \prod_{i=1}^{n} f(x_i, y_i) = \prod_{i=1}^{n} \underbrace{(1-p)^{y_i} p^{1-y_i}}_{\text{marginal density of } y_i} \phi(x_i, 0, 1)^{1-y_i} \phi(x_i, 1, 1)^{y_i}$$

here ϕ is the pdf of x_i with respect to $N(0,1)$ or $N(1,1)$ respectively

$$\log L(c) = \sum_{i=1}^{n} y_i \log(1-p) + (1-y_i)\log p + c_0$$

c_0 → doest involve p so can be considered some constant

$$E \log L(c)\big|_{x_1 \dots x_n} = \log(1-p) \sum_{i=1}^{n} E(y_i | x_1, \dots x_n) + \log p \sum 1 - E(y_i | x_1, \dots x_n) + E(c_0)$$

EM Algorithm

→ Start with $p^0 = 1 - q^0$ where $q^0 = \bar{X}$

E step → $E_{p^0}(\log L(c) | x_1 \dots x_n) = \log(1-p) \sum_{i=1}^{n} y_i^0 + \log p (n - \sum_{i=1}^{n} y_i^0)$
$+ \log C_0$

where $y_i^0 = E_{p^0}(y_i | x_1 \dots x_n) = E_{p^0}(y_i | x_i)$

Now

$$y_i = \begin{cases} 0 & \text{with probability } p \\ 1 & \text{" " } q \end{cases}$$

$E(y_i | x_1 \dots x_n) = E(y_i | x_i) = 0 \cdot P_i + 1(1 - P_i) = 1 - P_i = q_i$

conditional probability of $N(1,1)$

M-Step → Maximizing $E_{p^0}(\log L(c) | x_1 \dots x_n)$

gives

$$p^1 = \frac{n - \sum y_i^0}{n}$$ (by setting $\frac{d E_{p^0}(\log L(c) | x_1 \dots x_n)}{dp} = 0$)

Iteration → Repeat until convergence

\# $y_i^0 = E_{p^0}(y_i | x_i) = P(x_i \text{ belongs to } N(1,1) | x_i)$

(Bayes theorem) $= \dfrac{P(x_i | N(1,1)) P(N(1,1))}{P(x_i | N(1,1)) P(N(1,1)) + P(x_i | N(0,1)) P(N(0,1))}$

$= \dfrac{\phi(x_i, 1, 1)(1 - p^0)}{(1 - p^0)\phi(x_i, 1, 1) + p^0 \phi(x_i, 0, 1)}$

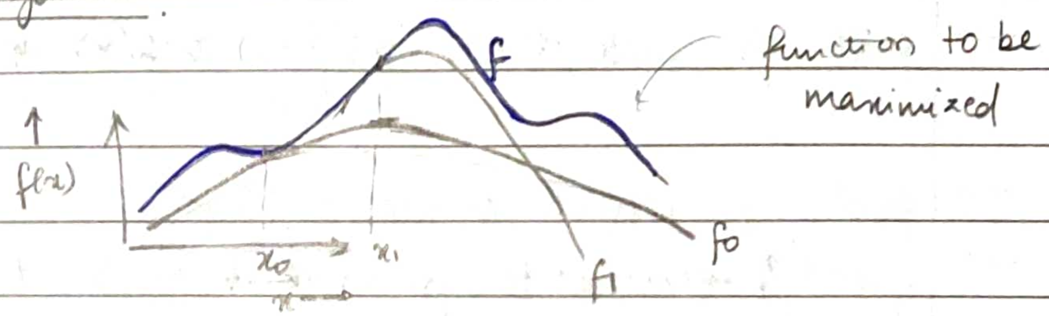
Page No.:

<u>Task</u>:- In the previous question if we have $N(\mu_1, \sigma_1^2)$ & $N(\mu_2, \sigma_2^2)$ where $\mu_1, \sigma_1, \mu_2, \sigma_2$ are also unknown along with p. Find MLE of p using E-M Algorithm

—— x ——

NEW MEETING
<u>26/06/2021</u>.

<u>MM Algorithm</u>

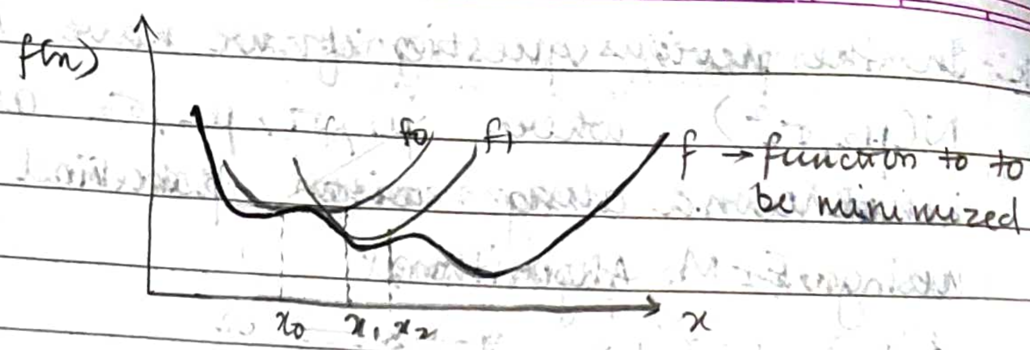


<u>Algorithm (Maximization)</u> [Minorization - Maximization Algorithm]

- Start with an initial x_0
- find f_0 such that $\begin{cases} f_0(x) \leq f(x) \ \forall x \\ f_0(x_0) = f(x_0) \end{cases}$
- find x_1 that maximizes $f_0(x)$
- define $f_1(x)$ such that $\begin{cases} f_1(x) \leq f(x) \ \forall x \\ f_1(x_1) = f(x_1) \end{cases}$

repeat the steps until convergence

f_0 is called <u>minorization function</u>



Algorithm - Minimization [Majorization Minimization Algorithm]

→ Start with x_0

→ find f_0 such that $\begin{cases} f_0(x) \geq f(x) \ \forall x \\ f_0(x_0) = f(x_0) \end{cases}$

→ find x_1 which minimizes f_0

→ find f_1 similar to f_0 and repeat till convergence

f_0 is called majorization function

Minorization - Maximization ⟶ Majorization - Minimization

f (maximization) ⟶ $-f$ (minimization)

→ Converting maximization function to minimization form, hence proving both the MM forms are actually the same.

— × —

Convergence of MM algorithm

Consider a maximization problem.
we need to prove that
$f(x_0) \leq f(x_1) \leq f(x_2) \leq \ldots\ldots$

If we can prove $f(x_1) \geq f(x_0)$, similarly we can say that the whole series is increasing

now,
we know that
$$f_0(x) \leq f(x) \quad \forall x \quad \& \longrightarrow (1)$$
$$f_0(x_0) = f(x_0) \longrightarrow (2)$$

now
$$\max_x \left(f_0(x) \right) = f_0(x_1)$$

$$\therefore f_0(x_0) \leq f_0(x_1) \Rightarrow f(x_0) \leq f_0(x_1) \ (\text{from } (2))$$
$$\Rightarrow f(x_0) \leq f_0(x_1) \leq f(x_1) \quad (\text{from } (1))$$
$$\Rightarrow f(x_0) \leq f(x_1)$$
$$\Rightarrow f(x_1) \geq f(x_0)$$

Similarly, we can prove that $f(x_{n+1}) \geq f(x_n)$.
Thus the series is increasing, so the convergence is guaranteed.

Similarly minimization problem convergence can be proved

Note:- If function is nice then convergence is guaranteed to the global maxima/minima acc. to the problem.

But, if there are many local maxima/minima, MM algorithm can get stuck in one of the local optima.

To solve this problem, the algo. has to be run several times with different initial values or other methods like simulated annealing / genetic algo's have to be used.

— x —

Task

Show that EM algorithm can be viewed as a special case of MM algorithm.

Proving convergence of IRWLS for LAD regression

In order to prove the convergence of IRWLS, it is enough to show that IRWLS for LAD regression can be viewed as a special case of of MM algorithm.

- Actual function to minimize → $\sum |y_i - \alpha - \beta x_i|$
- IRWLS ⟶ α_0, β_0 (start with random values)

function minimized → $\sum \frac{(y_i - \alpha - \beta x_i)^2}{|y_i - \alpha_0 - \beta_0 x_i|}$ → α_1, β_1 obtained

Repeat until convergence.

Page No.:

Let us define $m_i = \frac{|y_i - \alpha - \beta x_i|}{|y_i - \alpha_0 - \beta_0 x_i|}$ $\forall i = 1, 2 \ldots n$

$$m_i \geq 0 \quad \forall i$$

we know that

$$m_i \leq \frac{1}{2}(m_i^2 + 1) \quad \left(\because (m_i - 1)^2 \geq 0\right)$$

$\hookrightarrow$ (1)

Actual funⁿ: $\sum |y_i - \alpha - \beta x_i| = \sum m_i |y_i - \alpha_0 - \beta_0 x_i|$

Consider the funⁿ $\sum \frac{(m_i^2 + 1)}{2} |y_i - \alpha_0 - \beta_0 x_i|$

Minimization function

$$f(\alpha, \beta) = \sum |y_i - \alpha - \beta x_i| = \sum m_i |y_i - \alpha_0 - \beta_0 x_i|$$

Majorization function

$$f_0(\alpha, \beta) = \sum \frac{1}{2}(m_i^2 + 1) |y_i - \alpha_0 - \beta_0 x_i|$$

$\hookrightarrow \Big[f(\alpha, \beta) \leq f_0(\alpha, \beta) \quad \forall \alpha, \beta \quad$ (from (1)) &

$f(\alpha_0, \beta_0) = f_0(\alpha_0, \beta_0)$ as $m_i\big|_{\alpha_0, \beta_0} = 1$

& $m_i\big|_{\alpha_0, \beta_0} = \frac{1}{2}(m_i^2 + 1)\big|_{\alpha_0, \beta_0} \Big]$

Consider,

$$f_0(\alpha, \beta) = \sum \frac{1}{2}(m_i^2+1)\,|y_i - \alpha_0 - \beta_0 x_i|$$

$$= \sum \frac{1}{2}\left[\frac{(y_i-\alpha-\beta x_i)^2}{(|y_i-\alpha_0-\beta_0 x_i|)^2}+1\right]|y_i-\alpha_0-\beta_0 x_i|$$

$$= \underbrace{\frac{1}{2}}_{\text{constant}} \underbrace{\sum \frac{(y_i-\alpha-\beta x_i)^2}{|y_i-\alpha_0-\beta_0 x_i|}}_{\text{IRWLS for LAD regression}} + \underbrace{\frac{1}{2}\sum |y_i-\alpha_0-\beta_0 x_i|}_{\text{Independent of } \alpha, \beta}$$

$\therefore$ minimization of $f_0(\alpha,\beta) \equiv$ minimization of $\sum \frac{(y_i-\alpha-\beta x_i)^2}{|y_i-\alpha_0-\beta_0 x_i|}$

Hence we proved that IRWLS for LAD regression is a special case of MM algorithm.

As MM algorithm convergence is guaranteed, IRWLS for LAD regression " " also ".

—x—

Page No.:

<u>Task.</u>

Develop a ranking system for teams when there are matches 1v1 using all the methods learnt so far.

<u>Hint</u>:- Use Bradley - Terry Model

<u>For eg</u>:- There are 1, 2 ... 10 teams.
Let their capacity be some $\tau_1, \tau_2 \cdots \tau_{10}$

If Team i plays vs Team j,

$$P(i \text{ wins over } j) = \frac{\tau_i}{\tau_i + \tau_j}$$

Now,
we might have observations like

	VS	Wins
Team 1	Team 3	3 - 1
" 6	" 10	2 - 1
" 5	" 8	0 - 3
⋮	⋮	⋮

$$\text{Likelihood function} = \left[\binom{4}{1}\left(\frac{\tau_3}{\tau_3+\tau_1}\right)^1\left(\frac{\tau_1}{\tau_3+\tau_1}\right)^3\right] \cdot \left[\binom{3}{1}\left(\frac{\tau_{10}}{\tau_6+\tau_{10}}\right)\left(\frac{\tau_6}{\tau_6+\tau_{10}}\right)^2\right]$$

[multiplication because all events are independent]

Page No.:

problem with capacities τ

τ are not unique. Any set $k(\tau_1, \ldots, \tau_{10})$ are valid where k is some constant.

Hence, we need a starting assumption like $\tau_1 = 1$, on the basis of which we can find $\tau_2, \ldots, \tau_{10}$ in a relative sense.

—x—

Convergence of EM algorithm (general case)

Algorithm (EM)

→ Consider initial values of parameters. Let it be a vector θ_0

E step → Calculate $E_{\theta_0}(\underbrace{\log L_c(\theta)}_{\text{complete log likelihood of parameters } \theta} \mid \text{given data})$

- Let the given data be a vector $\boxed{X}$
- " " missing " " " " $\boxed{Y}$

now

$$E_{\theta_0}(\log L_c(\theta) \mid x) = \int \log L_c(\theta)\, \underbrace{f_{\theta_0}(y|x)}_{\text{conditional density}}\, dy = Q(\theta, \theta_0)$$

$\therefore$ E step :- Calculate $Q(\theta, \theta_0)$

M step → Maximize $Q(\theta, \theta_0)$ wrt θ to get θ_1

<u>Actual problem</u> :- Maximize $L(\theta)$ or $\log L(\theta)$

$\downarrow$

likelihood function using just the given data

Now, we know that

$$g_\theta(x, y) = f_\theta(x) \; h_\theta(y|x)$$

$g_\theta(x,y)$ $\downarrow$ joint pdf/pmf

$f_\theta(x)$ $\downarrow$ marginal pdf/pmf

$h_\theta(y|x)$ $\downarrow$ conditional pdf/pmf

X : Given/original Data

Y : Missing Data

(X, y) : complete Data

$$L_c(\theta) = L(\theta) \; h_\theta(y|x)$$

applying log on both sides

$$\log L_c(\theta) = \log L(\theta) + \log h_\theta(y|x)$$

$$\log L(\theta) = \log L_c(\theta) - \log h_\theta(y|x)$$

applying Expectation on both sides

$$E(\log L(\theta)|x) = E(\log L_c(\theta)|x) - E(\log h_\theta(y|x)|x)$$

$\downarrow$ ($\because$ it is only dependent on x)

$$\log L(\theta) = E(\log L_c(\theta)|x) - E(\log h_\theta(y|x)|x)$$

Page No.:

Starting with θ_0 $\rightarrow$ acc. to EM algorithm

$$E_{\theta_0}(\log L_c(\theta)|x) = Q(\theta, \theta_0) = \int \log(L_c(\theta)) h_{\theta_0}(y|x)\,dy$$

$$E_{\theta_0}(\log h_\theta(y|x)|x) = H(\theta, \theta_0) = \int \log(h_\theta(y|x))\, h_{\theta_0}(y|x)\,dy$$

$$\boxed{\log L(\theta) = Q(\theta, \theta_0) - H(\theta, \theta_0)}$$

Usual MLE estimation $\longrightarrow$ maximize $\log(L(\theta))$

EM Algo $\rightarrow$ Start with θ_0

Compute $Q(\theta, \theta_0) \xrightarrow{\text{maximize}}$ get θ_1

$Q(\theta, \theta_1) \xrightarrow{\quad " \quad}$ " θ_2

(E step) (m-step)

To prove convergence, it is necessary to show that $\log L(\theta_0) \leq \log L(\theta_1) \leq \log L(\theta_2) \ldots$

If we show $\log L(\theta_0) \leq \log L(\theta_1)$, others can be proved similarly

i.e.

we need to show

$$\log L(\theta_0) = Q(\theta_0, \theta_0) - H(\theta_0, \theta_0) \leq \log L(\theta_1) = Q(\theta_1, \theta_0) - H(\theta_1, \theta_0)$$

Now,

θ_1 is found by maximizing $Q(\theta, \theta_1)$

$\therefore$ $Q(\theta_0, \theta_0) \leq Q(\theta_1, \theta_0)$

We now need to show

$$H(\theta_1, \theta_0) \leq H(\theta_0, \theta_0)$$

$$H(\theta_1, \theta_0) = \int \log(h_{\theta_1}(y|x)) \; h_{\theta_0}(y|x)\, dy = E_{\theta_0}(\log h_{\theta_1}(y|x)|x)$$

$$H(\theta_0, \theta_0) = \int \log(h_{\theta_0}(y|x)) \cdot h_{\theta_0}(y|x)\, dy$$

$$H(\theta_1, \theta_0) - H(\theta_0, \theta_0) = \int \log\left[\frac{h_{\theta_1}(y|x)}{h_{\theta_0}(y|x)}\right] h_{\theta_0}(y|x)\, dy$$

Acc. to Jensen's inequality :-

→ If ϕ is convex then $E(\phi(x)) \geq \phi(E(x))$

Eg $\phi(x) = x^2$

→ If ϕ is concave then $E(\phi(x)) \leq \phi(E(x))$

→ If ϕ is linear " " = "

Now, log is is concave function,

$$\therefore H(\theta_1, \theta_0) - H(\theta_0, \theta_0) = E_{\theta_0}\left(\log\left[\frac{h_{\theta_1}(y|x)}{h_{\theta_0}(y|x)}\right] \Big| x\right)$$

$$\leq \log\left(E_{\theta_0}\left(\left[\frac{h_{\theta_1}(y|x)}{h_{\theta_0}(y|x)}\right] \Big| x\right)\right)$$

$$= \log \int \frac{h_{\theta_1}(y|x)}{\cancel{h_{\theta_0}(y|x)}} \cdot \cancel{h_{\theta_0}(y|x)}\, dy$$

$H(\theta_1, \theta_0) - H(\theta_0, \theta_0) \leq \log \int h_{\theta_1}(y|x)dy = \log 1 = 0$

$\therefore H(\theta_1, \theta_0) \leq H(\theta_0, \theta_0)$

Hence, $\log L(\theta_0) \leq \log L(\theta_1)$

Thus, convergence is guaranteed

Note: similar to MM algo, EM converges to global maxima if function is nice, else it might get stuck in any local minima.
Solution is re-running with different initial values or using simulated annealing or genetic algorithms.

~~NEM~~ NEW MEETING

3rd July 2021

Q. There are n teams, we need to develop a ranking system for them when we are given the information about 1 v 1 matches.

→ Suppose we have the information

Teams			Win ratio
1	vs	2	3 - 2
2	vs	3	0 - 2
3	vs	1	2 - 1

Let Team 1 → p_1
Team 2 → p_2
Team 3 → p_3

p_1, p_2, p_3 → capacities

Bradley Terry Model

Page No.:

$$\text{Likelihood} = L = C_0\left(\frac{p_1}{p_1+p_2}\right)^3\left(\frac{p_2}{p_1+p_2}\right)^2\left(\frac{p_3}{p_2+p_3}\right)^2\left(\frac{p_3}{p_1+p_3}\right)^2\left(\frac{p_1}{p_1+p_3}\right)$$

$$\log L = C_1 + 3\{\log p_1 - \log(p_1+p_2)\} + 2\{\log p_2 - \log(p_1+p_2)\}$$
$$+ 2\{\log p_3 - \log(p_2+p_3)\} + 2\{\log p_3 - \log(p_1+p_3)\}$$
$$+ \{\log p_1 - \log(p_1+p_3)\}$$

need to be separated for easier differentiation

Now

p_1, p_2, p_3 & kp_1, kp_2, kp_3 will lead to the same answer. Hence we set $p_1 = 1$

Now,

$$\log L = C_1 + 2\log p_2 + 4\log p_3 - 5\log(1+p_2) - 3\log(1+p_3) - 2\log(p_2+p_3)$$

Setting $\frac{\partial \log L}{\partial p_2} = \frac{\partial \log L}{\partial p_3} = 0$ does not give a direct answer

↳ we would have to use Newton-Raphson on

$$F = \frac{\partial \log L}{\partial p_2} = \frac{2}{p_2} - \frac{5}{1+p_2} - \frac{2}{p_2+p_3} = 0$$

$$G = \frac{\partial \log L}{\partial p_3} = \frac{4}{p_3} - \frac{3}{1+p_3} - \frac{2}{p_2+p_3} = 0$$

Now, we want to maximize log L but we can't do that directly. We use MM algorithm.

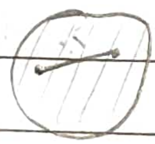
We now need a Minorization function which is easyly tracrable (Easy to find MAXIMA)

⇒ convex function

properties (necessary & sufficient)
$[f(\alpha x+\beta y) \leq \alpha f(x)+\beta f(y)]$
$f''(x) \geq 0, \alpha, \beta \geq 0, \alpha+\beta=1$

line joining any two pts in the graph lies inside the graph

⇒ convex set



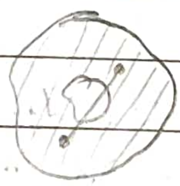
a set, in which line joining any two pts lies inside the set

⇒ concave function

properties (necessary & sufficient)
$[f(\alpha x+\beta y) \geq \alpha f(x)+\beta f(y)]$
$f''(x) \leq 0, \alpha, \beta \geq 0, \alpha+\beta=1$

condition for convex function is not always satisfied here

⇒ concave set



condition for convex set is not always satisfied here

Now, log is a concave function

$$\log(p_2+p_3) = \log\left(\alpha \frac{p_2}{\alpha} + \beta \frac{p_3}{\beta}\right) \geq \alpha \log\left(\frac{p_2}{\alpha}\right) + \beta \log\left(\frac{p_3}{\beta}\right)$$

We need α & β such that at initial point we have RHS = LHS & at all other pts original function ≥ surrogate. Also $\alpha+\beta=1$ must be satisfied.

Page No.:

Let our initial ~~et~~ choice be $p_1 = 1,\ p_2 = p_2^0,\ p_3 = p_3^0$

$$\log(p_2+p_3) \geq \frac{p_2^0}{p_2^0+p_3^0} \log\left(p_2 \cdot \frac{p_2^0+p_3^0}{p_2^0}\right) + \frac{p_3^0}{p_2^0+p_3^0} \log\left(p_3 \cdot \frac{p_2^0+p_3^0}{p_3^0}\right)$$

$$\Rightarrow \frac{p_2^0}{p_2^0+p_3^0} + \frac{p_3^0}{p_2^0+p_3^0} = 1 \ \& \ \text{both} > 0$$

$$\Rightarrow LHS\Big|_{p_2^0, p_3^0} = RHS\Big|_{p_2^0, p_3^0}$$

$$\Rightarrow LHS \geq RHS \quad \text{for } \forall (p_2, p_3)$$

$\therefore$ RHS is the minorization of $\log(p_2+p_3)$ at p_2^0, p_3^0

But we want minorization of $-\log(p_2+p_3)$
and $-$RHS is a majorization of $-\log(p_2+p_3)$.

$\therefore$ we need to find some function g such that $g(p_2, p_3) \leq -\log(p_2+p_3)$ &

$$g(p_2, p_3)\Big|_{p_2^0, p_3^0} = -\log(p_2+p_3)\Big|_{p_2^0, p_3^0}$$

Let us try fourier transform $\Rightarrow$ $f(x) = f(x_0) + (x-x_0) f'(x_0) + \frac{(x-x_0)^2}{2!} f''(\xi)$

If f is convex $f(x) \geq f(x_0) + (x-x_0) f'(x)$ $\left[\text{as } (x-x_0)^2 f''(\xi) \geq 0\right]$
" " " concave " $\leq$ " + " [" " " ≤ 0]

Page No.:

As log is concave

$\log(x) \leq \log(x_0) + (x - x_0)\, 1/x_0$

$-\log(x) \geq -\log(x_0) - (x - x_0)\, 1/x_0$

let $x = p_2 + p_3$, $x_0 = p_2^0 + p_3^0$

$$-\log(p_2+p_3) \geq -\log(p_2^0+p_3^0) - \frac{(p_2+p_3-p_2^0-p_3^0)}{p_2^0+p_3^0}$$

$\therefore$

$$\log L = C_1 + 2\log p_2 + 4\log p_3 - 5\log(1+p_2) - 3\log(1+p_3) - 2\log(p_2+p_3)$$

$$\geq C_1 + 2\log p_2 + 4\log p_3 - 5\left\{\log(1+p_2^0) + \frac{(p_2-p_2^0)}{1+p_2^0}\right\} - 3\left\{\log(1+p_3^0) + \frac{(p_3-p_3^0)}{1+p_3^0}\right\} - 2\left\{\log(p_2^0+p_3^0) + \frac{(p_2+p_3-p_2^0-p_3^0)}{p_2^0+p_3^0}\right\}$$

= minorization of $\log L$

$= (\log L)^0$

Algorithm:-

→ Start with $p_1 = 1$, $p_2 = p_2^0$, $p_3 = p_3^0$

→ Set minorization of $\log(L)$ using taylor's series as

→ Maximize the minorization $(\log L)^0$.

→ $\dfrac{\partial(\log L)^0}{\partial p_2} = \dfrac{2}{p_2} - \dfrac{5}{1+p_2^0} - \dfrac{2}{p_2^0+p_3^0}$ | $\dfrac{\partial(\log L)^0}{\partial p_3} = \dfrac{4}{p_3} - \dfrac{3}{1+p_3^0} - \dfrac{2}{p_2^0+p_3^0}$

we set $\dfrac{\partial(\log L)^0}{\partial p_2} = \dfrac{\partial(\log L)^0}{\partial p_3} = 0$ to get p_2^1, p_3^1

→ Repeat until convergence

Bradley Terry Model Incorporating Home Advantage

→ Let θ be a parameter of Home Advantage

p_A be the capacity of team A &
p_B " " " " team B] Bradley terry parameters

→ P(A wins over B, when played at Home) $= \dfrac{p_A+\theta}{p_A+p_B+\theta}$

P(" " " " , " " abroad) $= \dfrac{p_A}{p_B+p_A+\theta}$

→ let us assume we have information like before

Home Abroad	Win Ratio
1 vs 2	3 - 2
2 vs 3	0 - 2
3 vs 1	2 - 1

$$L = c_0 \left(\frac{p_1+\theta}{p_1+p_2+\theta}\right)^3 \left(\frac{p_2}{p_1+p_2+\theta}\right)^2 \cdots$$

$$\log L = c_1 + 3\log(p_1+\theta) + 2\log p_2 - 5\log(p_1+p_2+\theta)$$

we need a minorization of log L

using concave function property
$$\log(p_1+\theta) \geq \frac{p_1^0}{p_1^0+\theta^0}\log\left(p_1 \cdot \frac{p_1^0+\theta^0}{p_1^0}\right) + \frac{\theta^0}{p_1^0+\theta^0}\log\left(\theta \cdot \frac{p_1^0+\theta^0}{\theta^0}\right)$$

using Taylor's series
$$-\log(p_1+p_2+\theta) \geq -\log(p_1^0+p_2^0+\theta^0) - \frac{(p_1+p_2+\theta-p_1^0-p_2^0-\theta^0)}{p_1^0+p_2^0+\theta^0}$$

We can construct a minorization function using the above inequalities for all cases.

The we can set $p_1 = k$ = constant and iterative determine $p_2, p_3 \dots$ using MM algorithm by maximizing the minorization function in each step.

Note: In case some information is missing, we ignore that and construct the likelihood with only the known information unlike in EM algorithm.

Case of Race (Modified Bradley Terry Model)

eg

	A	B	C	D	E	→ participants
Race 1 →	B	C	A	–	–	
Race 2 →	C	D	E	B	A	
	1st	2nd	3rd	4th	5th	

⟶ Rankings in 2 races

$$\mathcal{L} = C_0 \underbrace{\left(\frac{p_B}{p_C + p_B + p_A} \cdot \frac{p_C}{p_C + p_A} \right)}_{\text{Race 1}} \cdot \underbrace{\left(\frac{p_C}{p_C + p_A + p_E + p_B + p_D} \cdot \frac{p_D}{p_D + p_E + p_B + p_A} \cdot \frac{p_E}{p_E + p_B + p_A} \cdot \frac{p_B}{p_B + p_A} \right)}_{\text{Race 2}}$$

We use the method use before (MM Algorithm) to rank these players.

Case of Inteviewee's with a combination of a panel of Interviewers

Interviewees →	A	B	C	D	E
Interviewee 1 →	D	Z	X	X	X
" 2 →	B	C	A	X	X
" 3 →	D	A	C	E	X
	1st	2nd	3rd	4th	5th

↳ Ranking Acc. to Various Interviees

The problem is similar to the race problem. So it can be solved in the same way

Alternate Solution Suggested by Me for case of Race.

	1st	2nd	3rd	4th	5th
Race 1	B	C	A	X	X
Race 2 →	C	D	E	B	A

$$L = C_0\left(\frac{p_B}{p_B+p_C}\cdot\frac{p_B}{p_B+p_A}\cdot\frac{p_C}{p_C+p_A}\right)\cdot\left(\frac{p_C}{p_C+p_D}\cdot\frac{p_C}{p_C+p_E}\cdot\frac{p_C}{p_C+p_B}\cdot\frac{p_C}{p_C+p_A}\right.$$

$$\cdot\frac{p_D}{p_D+p_E}\cdot\frac{p_D}{p_D+p_B}\cdot\frac{p_D}{p_D+p_A}\cdot$$

$$\left.\frac{p_E}{p_E+p_B}\cdot\frac{p_E}{p_E+p_A}\cdot\frac{p_B}{p_B+p_A}\right)$$

Page No.:

Difference b/w my approach & Bradley Terry model for races.

My Model :- For Race 1 = [B, C, A]
I consider (B vs C, B vs A) & (C vs A). Basically pairs of two, and then write the likehood accordingly

Bradly Terry :- For Race 1=[B, C, A],
B.T considers (B vs(C&A)), (C vs A)
Basically he divides the whole set as winning vs losing and writes the likelihood accordingly.

We need to code out and compare if these both approaches give the same result or one is better than the other.

I have also made a 1v1 model using EM Algorithm. That can be compared using sirs MM approach.

Proof that EM is special case of MM algorithm

X : Observed Data, Y: Missing Data, (X,Y): complete Data

θ : Parameters to be estimated

$$L_c(\theta) = \underbrace{g_\theta(x,y)}_{\text{joint pdf/pmf}} = \underbrace{f_\theta(x)}_{\text{marginal pdf/pmf} = L(\theta)} \; \underbrace{h_\theta(y|x)}_{\text{conditional pdf/pmf}}$$

$$\log L_c(\theta) = \log f_\theta(x) + \log h_\theta(y|x)$$

$$E(\log L_c(\theta) \mid X) = E(\log f_\theta(x) \mid X) + E(\log h_\theta(y|x) \mid X)$$

$$E(\log L_c(\theta) \mid X) = E(\log L(\theta) \mid X) + E(\log h_\theta(y|x) \mid X)$$

$$E(\log L_c(\theta) \mid X) = \log L(\theta) + E(\log h_\theta(y|x) \mid X)$$

(as $L(\theta)$ is likelihood only based on 'X')

$$\log L(\theta) = E(\log L_c(\theta) \mid X) - E(\log h_\theta(y|x) \mid X)$$

at initial values of parameters θ_0

$$\log L(\theta) = E_{\theta_0}(\log L_c(\theta) \mid X) - E_{\theta_0}(\log h_\theta(y|x) \mid X)$$

$$\boxed{\log L(\theta) = Q(\theta,\theta_0) - H(\theta,\theta_0)}$$

now, let

$$\log L_0(\theta) = Q(\theta,\theta_0) - H(\theta_0,\theta_0)$$

Thinking is difficult, that's why most people judge.

– Carl Jung

now,

we need to prove that

1. $\log L_0(\theta)$ is minorization of $\log L(\theta)$
2. Maximization of $\log L_0(\theta)$ is equal to maximization of $Q(\theta, \theta_0)$

$2 \rightarrow$ now,

$\log L_0(\theta) = Q(\theta, \theta_0) - H(\theta_0, \theta_0)$

$\hookrightarrow$ constant & free of θ

$\therefore \max\limits_{\theta}(Q(\theta, \theta_0)) = \max\limits_{\theta}(\log L_0(\theta))$

as $H(\theta_0, \theta_0)$ is constant

point 2 proved.

$1 \rightarrow$ We already know that

$\log L(\theta_0) = Q(\theta_0, \theta_0) - H(\theta_0, \theta_0) = \log L_0(\theta_0)$

We need to show that

$\log L(\theta) \geq \log L_0(\theta) \quad \forall \theta$

i.e $\log L(\theta) - \log L_0(\theta) = H(\theta_0, \theta_0) - H(\theta, \theta_0) \geq 0$

has to be proved

Now, we need to prove

$H(\theta_0, \theta_0) - H(\theta, \theta_0) \geq 0$ or

$H(\theta, \theta_0) - H(\theta_0, \theta_0) \leq 0$

$$H(\theta, \theta_0) - H(\theta_0, \theta_0) = \int \log\left(\frac{h_\theta(y|x)}{h_{\theta_0}(y|x)}\right) h_{\theta_0}(y|x)\, dy$$

As log is a concave function, jensens inequality states that $E(\phi(x)) \leq \phi(E(x))$ when ϕ is concave

$$\therefore H(\theta, \theta_0) - H(\theta_0, \theta_0) = E_{\theta_0}\left(\log\left(\frac{h_\theta(y|x)}{h_{\theta_0}(y|x)}\right) \Big| x\right)$$

$$\leq \log\left(E_{\theta_0}\left[\frac{h_\theta(y|x)}{h_{\theta_0}(y|x)}\right] \Big| x\right)$$

$$= \log \int \frac{h_\theta(y|x) \cdot \cancel{h_{\theta_0}(y|x)}}{\cancel{h_{\theta_0}(y|x)}}\, dy$$

$$= \log \int h_\theta(y|x)\, dy = \log 1 = 0$$

$H(\theta, \theta_0) - H(\theta_0, \theta_0) \leq 0$

$\therefore$ point 1 proved.

Thus, $\log L_0(\theta)$ is minorization of $\log L(\theta)$ & maximization of $\log L_0(\theta) \equiv$ maximization of $Q(\theta, \theta_0)$

Hence we can say that EM algorithm is a special case of MM algorithm.

If you haven't the strength to impose your own terms upon life, you must accept the terms it offers you.

– T. S. Eliot

Q) Estimating the lifetime of a bulb when we just have information of a sample for a fixed amount of time

⇒ Let us suppose that an electric bulb follows exponential distribution with mean θ

$\therefore f_\theta(x) = \frac{1}{\theta} e^{-x/\theta}, \ x>0 \rightarrow$ pdf

i) Given info :- n bulbs are switched on after t hours m have stopped working

Solution :- total bulbs $\rightarrow n$
bulbs working after 't' hours $\rightarrow n-m$
" stopped working before 't' hrs $\rightarrow m$

P(A bulb works after 't' hrs) $= \int_t^\infty f_\theta(x)\,dx = \int_t^\infty \frac{1}{\theta} e^{-x/\theta} d\theta$
$= e^{-t/\theta}$

P(A stops befor 't' hrs) $= 1 - e^{-t/\theta}$

$L(\theta) = \binom{n}{m} \left(1-e^{-t/\theta}\right)^m \left(e^{-t/\theta}\right)^{n-m} \rightarrow$ similar to Bin(n,p) where $p = 1-e^{-t/\theta}$

$L(\theta) = \binom{n}{m} (p)^m (1-p)^{n-m}$

MLE of $p = \hat{p} = \frac{m}{n}$

Now, $1-p = e^{-t/\theta} \Rightarrow [\log(1-p)] = \frac{-t}{\theta}$
$\Rightarrow \theta = \frac{-t}{\log(1-p)}$

$\therefore \hat{\theta} = \frac{-t}{\log(1-\hat{p})} = \frac{-t}{\log\left(\frac{n-m}{n}\right)}$

When little people are overwhelmed by big emotions, it's our job to share our calm, not join their chaos.
– L. R. Knost

→ found earlier in balls problem

$\therefore$ $\boxed{\text{MLE of } \theta = \hat{\theta} = \dfrac{-t}{\log\left(\frac{n-m}{n}\right)}}$

using if MLE of $\theta = \hat{\theta}$ then MLE of $f(\theta) = f(\hat{\theta})$

ii) Given info :- we have the stopping times of each of the bulbs. let it be $x_1, x_2 \ldots x_n$

Solution :- $L = \prod_{i=1}^{n} \frac{1}{\theta} e^{-x_i/\theta} \Rightarrow \log L = \log\left(\frac{1}{\theta^n} e^{-\Sigma x_i/\theta}\right)$

$$\log L = -n \log \theta - \frac{\Sigma x_i}{\theta}$$

$\frac{\partial \log L}{\partial \theta} = \frac{-n}{\theta} + \frac{\Sigma x_i}{\theta^2} \longrightarrow$ set it to zero

$$\text{MLE of } \theta = \hat{\theta} = \frac{\Sigma x_i}{n}$$

$$\boxed{\hat{\theta} = \bar{x}}$$

iii) Given info :- We have 'n' bulbs observed for 't' hrs 'm' have stopped working and 'n-m' are still working at 't' hours. For the 'm' non-working bulbs, we have the stopping time

$$\underbrace{x_1, x_2, \ldots\ldots x_m}_{\substack{\text{stopping time} \\ \text{of m bulbs} \\ \text{(known}}}, \underbrace{x_{m+1} \ldots\ldots x_n}_{\substack{\text{still working at 't' hours} \\ \text{stopping time unknown}}}$$

The greatest prison people live in is the fear of what other people think.

– David Icke

Solution :- • known data :- $x_1, x_2, \ldots\ldots x_m$, t
(stopping times of m bulbs)
• unknown data :- $x_{m+1}, x_{m+2} \ldots\ldots x_n$
(stopping times of 'n-m' bulbs that were working at time = 't' hrs)
• Also known that 'n-m' bulbs have life $\geq t$ hrs

complete likelihood $\rightarrow L_c = \prod_{i=1}^{m} \frac{1}{\theta} e^{-x_i/\theta} \cdot \prod_{i=m+1}^{n} \frac{1}{\theta} e^{-x_i/\theta}$

$$\log L_c = \log\left(\frac{1}{\theta^n} e^{-\Sigma x_i/\theta}\right) = -n\log\theta - \sum_{i=1}^{n} \frac{x_i}{\theta}$$

$$\boxed{\log L_c = -n\log\theta - \sum_{i=1}^{n} \frac{x_i}{\theta}}$$

EM Algorithm

$\rightarrow$ let us set $\theta = \theta_0$ with $\theta_0 = {-t}/\log\left(\frac{n-m}{n}\right)$ [found in part 1]

$\rightarrow$ E step

$$E(\log L_c \mid x_1 \ldots x_m, t) = E\left(-n\log\theta - \sum_{i=1}^{n} \frac{x_i}{\theta} \,\middle|\, x_1 \ldots x_m, t\right)$$

$$= -n\log\theta - \sum_{i=1}^{m} \frac{x_i}{\theta} - \sum_{i=m+1}^{n} E(x_i \mid x_1 \ldots x_m,$$

$$E_{\theta_0}(\log L_c \mid x_1 \ldots x_m, t) = -n\log\theta - \sum_{i=1}^{m} \frac{x_i}{\theta} - \sum_{i=m+1}^{n} \frac{x_i^0}{\theta}$$

where $x_i^0 = E_{\theta_0}(x_i \mid x_1 \ldots x_m, t)$, $i \in \{m+1, \ldots n\}$

We know that,

$x_i > t \quad \forall\, i \in \{m+1, m+2 \ldots\ldots n\}$

$x_i \sim$ exponential $(t+\theta_0)$ (using memoryless property)

$$E_{\theta_0}(x_i \mid \text{given data}) = t+\theta_0$$

$$E_{\theta_0}(\log L_c \mid x_1, \ldots x_m, t) = -n\log\theta - \sum_{i=1}^{m}\frac{x_i}{\theta} - (n-m)\frac{(\theta_0+t)}{\theta}$$

→ M step: We need to maximize $E_{\theta_0}(\log L_c \mid x_1 \ldots x_m, t)$ wrt θ.

$$\frac{d\, E_{\theta_0}(\log L_c \mid \text{given data})}{d\theta} = -\frac{n}{\theta} + \sum_{i=1}^{m}\frac{x_i}{\theta^2} + (n-m)\frac{(\theta_0+t)}{\theta^2}$$

Set the above eqn to zero

$$\frac{(n-m)(\theta_0+t)+\sum_{i=1}^{m} x_i}{\theta^{\cancel{2}}} = \frac{+n}{\cancel{\theta}}$$

$$\boxed{\theta_1 = \frac{(n-m)(\theta_0+t)+\sum_{i=1}^{m} x_i}{n}}$$

→ Repeat until convergence

Convergence

$$\theta = \phi(\theta) = \frac{n-m(\theta-t)+\sum x_i}{n}$$

$$|\phi'(\theta)| = \left|\frac{n-m}{n}\right| < 1, \text{ hence converges}$$

Notes

Convergence to the Right Value

$$\theta_k = \phi(\theta) = \frac{(n-m)(\theta_{k-1}+t) + \sum_{i=1}^{m} x_i}{n}$$

Applying $\lim k \to \infty$ on both sides
let $\theta_k \to \theta$ at $k \to \infty$

$$\lim_{k\to\infty} \theta_k = \lim_{k\to\infty} \frac{n-m}{n}(\theta_{k-1}+t) + \sum_{i=1}^{m} \frac{x_i}{n}$$

$$\theta = \frac{n\theta - m\theta}{n} + \frac{(n-m)t}{n} + \sum_{i=1}^{m} \frac{x_i}{n}$$

$$\cancel{n\theta} = \cancel{n\theta} - m\theta + (n-m)t + \sum_{i=1}^{m} x_i$$

$$\boxed{\theta = \frac{(n-m)t + \sum_{i=1}^{m} x_i}{m}} \longrightarrow \text{Actual MLE of } \theta$$

Hence, converges to ~~the~~ correct value

July 10 2021

Discussions
- Discussed the Bulb problem & Race code
- Discussed the Depth of Data & Median as an estimator

Task
- Asked to code out better for the race
- Asked to complete the report ASAP