1(b) Inverse transform method

(1) Let $U \sim U(0,1)$

$$X = -\log(U)$$

$$p = P(X > 1)$$

$$T_m = \frac{1}{m}\sum_{i=1}^{m} \mathbb{1}(X_i > 1) = p = e^{-1}$$

Hence value of e can be estimated

2(a) Extreme value distribution

$U_1, U_2, \ldots U_n \sim U(0,1)$

$S_n = \sum_{i=1}^{n} U_i \quad n \geq 1$

$T = \inf\{n : S_n > 1\}$, then $E(T) = e$

Proof:–

$$P(T=n) = P(S_{n-1} < 1 \text{ and } S_n > 1)$$

$$= P(S_{n-1} < 1) - P(S_n < 1)$$

$$= \frac{1}{(n-1)!} - \frac{1}{n!} = \frac{n-1}{n!} \quad n = 2, 3, \ldots$$

$$E(T) = \sum_{n \geq 2}^{\infty} \frac{n(n-1)}{n!} = \sum_{m \geq 0} \frac{1}{m!} = e$$

2(b) EVD with antithetic variables

We seek two unbiased estimators. Y' & Y'' for parameter I, having strong negative correlation.

$$\therefore E\left[\tfrac{1}{2}(Y' + Y'')\right] = I$$

$$Var\left[\tfrac{1}{2}(Y' + Y'')\right] = \tfrac{1}{4} Var(Y') + \frac{1}{4}(Var(Y'')) + \frac{1}{2} Cov(Y', Y'')$$

$$\boxed{Y' = \sum_{i=1}^{n} U_i \text{ \& } Y'' = \sum_{i=1}^{n} (1 - U_i)}$$

(Now proceed as above)

3.) Bootstrap method :-

The principle behid bootstrap method is given by:-
Out of N objects (distinct), the probability of a particular object being NOT selected is given by:-

$$p_1 = \left(1 - \frac{1}{N}\right)$$

If we perform N trials of this experiment, ~~and~~ the probability of it being not selected in any of the ~~experim~~ trials is given by

$$p_2 = \left(1 - \frac{1}{N}\right)^N$$

Now, as $N \to \infty$, $p_2 \to e^{-1}$, and hence acts as an approximation of e.

4.) Variant of EVD

$$U_1; U_2, \ldots U_n \sim U(0,1)$$

$$S = U_{(1)}, U_{(2)}, \ldots U_{(n)} \quad \text{[order statistics]}$$

$$M = \text{cumsum}(1/S)/n$$

i.e $$\boxed{M = \frac{1}{nU_{(1)}}, \frac{1}{nU_{(1)}} + \frac{1}{nU_{(2)}}, \ldots \frac{1}{n}\sum_{i=1}^{n} \frac{1}{U_{(i)}}}$$

$$t = \inf\{i : M_i > 1\}$$

Then $$\boxed{\frac{2}{S[t] + S[t-1]} \text{ is an estimator of } e}$$

5. Definition of pdf of $N(0,1)$

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$f(\sqrt{2}) = \frac{1}{\sqrt{2\pi}} e^{-1} \Rightarrow \boxed{\hat{e} = \frac{1}{f(\sqrt{2}) \times \sqrt{2\pi}}}$$

$$\text{A.} \because \boxed{f(x) = \frac{F(x+h) - F(x)}{h} \quad \lim_{h \to 0}}$$

$$\therefore f(\sqrt{2}) = \frac{F(\sqrt{2}+h) - F(\sqrt{2})}{h}$$

$$= \lim_{h \to 0} \frac{\Phi(X < \sqrt{2}) - \Phi(X < \sqrt{2}+h)}{h}$$

Hence, $\boxed{\hat{e} = \frac{1}{\sqrt{2\pi} * f(\sqrt{2})}}$

NUMERICAL APPROXIMATIONS :-

6. $$e^{-1} = \lim_{n \to \infty} \left(1 - \frac{1}{n}\right)^n$$

7. $$e = \frac{1}{2} \sum_{k=0}^{\infty} \frac{k+1}{k!}$$

8. $$e = \sum_{k=0}^{\infty} \frac{3 - 4k^2}{(2k+1)!}$$

Q-5 Let $L(Y^N;\theta)$ denote objective likelihood function.
After a cycle of EM step, we obtain θ_{t+1} from θ_t such that

$$L(Y^N;\theta_{t+1}) > L(Y^N;\theta_t)$$

By iterating EM steps, the algorithm should converge to MLE. The difference of parameters can be approximated as:-

$$\boxed{\theta_{t+1} \simeq \theta_t + G_X^{-1}(\theta_t)\,\partial L(Y^N;\theta_t)}$$ - EM update rule

where $\partial = (\partial_1, \cdots \partial_n)^T = \left(\partial/\partial\theta^1, \frac{\partial}{\partial\theta^2}, \cdots \frac{\partial}{\partial\theta^n}\right)^T$

$G_X(\theta) = (g_{X_{ij}}(\theta))$ is the Fisher information matrix of $p(x;\theta)$

$$\therefore g_{X_{ij}}(\theta) = -E_{p(x;\theta)}\left[\frac{\partial l(x;\theta)}{\partial_i\partial_j}\right]$$

Note 1: [$x = (y, z)$ where y = visible data, z = hidden data]

Now, Fisher's scoring update rule is

$$\boxed{\theta_{t+1} \simeq \theta_t + G_Y^{-1}(\theta_t)\,\partial L(Y^N;\theta_t)}$$ - Fisher update Rule

$G_Y(\theta)$ is the Fisher information matrix of $p(y;\theta)$ (visible data)

$$\therefore g_{Y_{ij}}(\theta) = -E_{p(y;\theta)}\left[\frac{\partial\, l(y;\theta)}{\partial_i\partial_j}\right]$$

Note 2 :- G_Y^{-1} is intractable is most problems, which is why EM algorithm is important.

Now, we know that

$$-l(y;\theta) = -l(x;\theta) + l(z|y;\theta)$$

$$-E_{p(y;\theta)}\left[\frac{\partial}{\partial_i\partial_j} l(y;\theta)\right] = -E_{p(x;\theta)}\left[\frac{\partial}{\partial_i\partial_j} l(x;\theta)\right] + E_{p(x;\theta)}\left[\frac{\partial\, l(z|y;\theta)}{\partial_i\partial_j}\right]$$

$\Rightarrow \quad G_Y(\theta) = G_X(\theta) - G_{Z|Y}(\theta)$

$G_{Z|Y}$ is a conditional Fisher Information matrix defined similar to its counterparts.

By simultaneous diagonalization of G_X, G_Y & $G_{Z|Y}$ we obtain:-

$$\boxed{G_Y^{-1} = \left(I + \sum_{i=1}^{\infty} (G_X^{-1} G_{Z|Y})^i\right) G_X^{-1}} \quad \text{---} \quad \#$$

Using above equation in Fisher update rule we obtain

$$\begin{aligned} \theta_{t+1} &= \theta_t + G_Y^{-1} \partial L(Y^N; \theta_t) \\ &= \theta_t + G_X^{-1} \partial L(Y^N; \theta_t) + G_X^{-1} G_{Z|Y} G_X^{-1} \partial L(Y^N, \theta_t) \\ &\quad + (G_X^{-1} G_{Z|Y})^2 G_X^{-1} \partial L(Y^N, \theta) \\ &\quad \ldots \text{ and so on} \end{aligned}$$

(★) shows that EM update is first order approximation of the Fisher's update rule.

Hence it can be attributed to the faster convergence rate of Fisher's scoring algorithm as compared to EM algorithm.

Simulation study performed testifies this statement