CSC 860 / MATH 795: Stochastic Optimization & Simulation Methodology
Fall 2022 Assignment 5

Mariia Aleksandrovych, Yaffa Atkins, Atoosa Lotfi, Olga Kliuchnik

Exercises 7.4, (7.7 and 8.11), 8.13, 8.20, 8.21.

EXERCISE 7.4. Consider a random variable $X(\theta)$. Assume that $\theta = \mathbb{E}[X(\theta)]$ is a scale parameter of the distribution F_θ; in other words, using the inverse function representation, $X(\theta) = F_\theta^{-1}(U) = \theta F_1^{-1}(U)$. Explain which of the following is a sufficient condition for:

$$\mathbb{E}\left[\frac{X(\theta)}{\theta}\right] = \frac{d}{d\theta}\mathbb{E}[X(\theta)].$$

(a) $\mathbb{E}[|X(1)|] < \infty$,

(b) $\mathbb{E}[X(\theta)^2] < \infty$, or

(c) $\mathbb{E}[|\theta X(\theta)|] < \infty$.

a) Expectation value of the X(theta) bounded is sufficient condition for the interchanging derivative and expectation of X(theta). However, when theta =1 it doesn't have any information about theta that is increase

b) $E(X(\theta)^2) = E\left(\theta^2\left(F_1^{-1}(U)\right)^2\right) = \theta E\left(\left(F_1^{-1}(U)\right)^2\right)$

This is one is not sufficient to satisfy the condition

c) $E\left(\theta X\left(\theta\right)\right) = \theta E\left(X\left(\theta\right)\right) < \infty \Rightarrow E\left(X\left(\theta\right)\right) < \infty$ which is sufficient for the condition above

EXERCISE 7.7. Repeat the stochastic approximation method for the mining investment problem of Example 7.3 using IPA, SF and MVD for derivative estimation. Assume here that the estimators are unbiased.

(a) Using simulation, estimate the corresponding confidence intervals and CPU times for the three derivative estimation methods for $\theta = 5, 8, 10, 12, 15$ and compare.

(b) Apply an appropriate theorem from Part I to establish convergence of the stochastic approximation to the true optimal value θ^*. Specify your choice of step size sequence and, for each method, verify the assumptions of the theorems that you use.

(c) Run the stochastic approximations and discuss.

Exersize 7.7.

$\boxed{IPA}$ $\quad X(\theta) \stackrel{\mathcal{L}}{=} -500\ln(1-u)$

$$\frac{dX(\theta)}{d\theta} = \frac{X(\theta)}{\theta} = -50\ln(1-u)$$

$$h = \frac{x}{\sqrt{x+1}} \qquad h'(X(\theta)) = \frac{1}{\sqrt{X+1}} - \frac{X}{2(1+X)^{3/2}}$$

Then algorithm,

$$\theta_{n+1} = \theta_n + \varepsilon\left(\frac{X(\theta_n)}{\theta_n} \cdot h'(X(\theta_n)) - 1\right) =$$

$$= \theta_n + \varepsilon\left(\frac{X(\theta_n)}{\theta_n (X(\theta_n)+1)^{\frac{1}{2}}} - \frac{X^2(\theta_n)}{2\theta_n (X(\theta_n)+1)^{3/2}} - 1\right)$$

$\boxed{\text{Score Function}}$

$$SF = \frac{\partial}{\partial\theta}\ln(f_\theta(X))$$

Then $\dfrac{d}{d\theta}\mathbb{E}[h(X(\theta))] = \mathbb{E}[h(X(\theta)) \cdot SF(\theta, X(\theta))]$

for exponential

$$f_\theta(x) = \exp\left(-\frac{x}{500}\right)\frac{1}{500}$$

$$SF = \frac{\partial}{\partial\theta}\ln\left(\frac{1}{50\theta} e^{-\frac{X(\theta)}{50\theta}}\right) =$$

$$= \frac{-\frac{1}{50\theta^2}\cancel{e^{-\frac{X(\theta)}{50\theta}}} + \cancel{\frac{1}{50\theta}}\cancel{e^{-\frac{X(\theta)}{50\theta}}} \cdot \left(+\frac{X(\theta)}{50\theta^2}\right)}{\cancel{\frac{1}{50\theta}}\cancel{e^{-X(\theta)/50\theta}}} =$$

$$= \left(-\frac{1}{\theta} + \frac{X(\theta)}{50\theta^2}\right) = \frac{1}{\theta}\left(\frac{X(\theta)}{50\theta} - 1\right)$$

Then algorith

$$\theta_{n+1} = \theta_n + \varepsilon\left[\frac{X(\theta_n)}{(X(\theta_n)+1)^{\frac{1}{2}}} \cdot \frac{1}{\theta_n}\left(\frac{X(\theta_n)}{50\theta_n} - 1\right) - 1\right]$$

for $X(\theta) \overset{\mathcal{L}}{\approx} -50\theta \ln(1-u)$

$\boxed{\text{MVD}}$ $\quad f_\theta = \frac{1}{50\theta} e^{-\frac{X(\theta)}{50\theta}}$

$$\frac{\partial}{\partial\theta} f_\theta(X) = -\frac{1}{50\theta^2} e^{-\frac{X(\theta)}{50\theta}} + \frac{1}{50\theta} e^{-\frac{X(\theta)}{50\theta}} \cdot \left(+\frac{X(\theta)}{50\theta^2}\right) =$$

$$= \frac{1}{\theta}\left(-\frac{1}{50\theta} e^{-\frac{X}{50\theta}} + \frac{1}{(50\theta)^2} e^{-\frac{X}{50\theta}}\right) =$$

$$= \frac{1}{\theta}\left(\frac{X - 50\theta}{(50\theta)^2}\right) e^{-\frac{X}{50\theta}}$$

$$\begin{cases} f^+ = \dfrac{X(\theta) - 50\theta}{(50\theta)^2} e^{-\frac{X}{50\theta}}, & X > 50\theta \\ f^- = -\dfrac{(X(\theta) - 50\theta)}{(50\theta)^2} e^{-\frac{X}{50\theta}}, & X \leq 50\theta \end{cases}$$

$c_\theta = \frac{1}{\theta}$

Then derivative

$$\frac{d}{d\theta}\mathbb{E}\left[h(X(\theta))\right] = \frac{1}{\theta}\mathbb{E}\left[h(X^+(\theta)) - h(X^-(\theta))\right]$$

where $X^\pm \overset{\mathcal{L}}{\approx} f_\theta^\pm \qquad h(X(\theta)) = \dfrac{X(\theta)}{\sqrt{1+X(\theta)}}$

Then algorithm

$$\theta_{n+1} = \theta_n + \varepsilon \left[\frac{1}{\theta_n} \mathbb{E}\left[h(X^+|\theta_n)) - h(X^-|\theta_n)) \right] - 1 \right]$$

$$X^{\pm} \overset{d}{\approx} f^{\pm}$$

b) applying Theorem 6.2. we have

The algorithm $\theta_{n+1} = \theta_n + \varepsilon_n Y_n$

Then G is bounded and continious, with unique point θ^* such that $G(\theta^*) = 0$ and $\hat{A} = \nabla G(\theta^*)$ is a Hurwitz with eigenvalue $\lambda_1, \dots \lambda_d, \dots$

$$B_n = \mathbb{E}\left[Y_n | \mathcal{F}_{n-1} \right] - G(\theta_n)$$

$$V_n = \mathbb{E}\left[(Y_n - \mathbb{E}\left[Y_n | \mathcal{F}_{n-1} \right])^2 \right]$$

there is the constant γ, β, δ such that

- $\varepsilon_n = n^{-\gamma}$, $\|B_n\| = O(n^{-\beta})$, $V_n = O(n^{-\delta})$
- $\gamma + \beta > 1$, $2\gamma + \delta > 1$

Then $\theta_n \to \theta^*$ w.p. 1., and if $\gamma < 1$

$$\mathbb{E}\left[\|\theta_n - \theta^*\|^2 \right] = O(n^{-\varkappa}), \quad \varkappa \overset{def}{=} \min(2\beta, \gamma+\delta)$$

if $\gamma = 1$, then above condition satisfied provided that $\lambda_{min} > \max(2\beta, 1+\delta, 1)$.

So, $X(\theta) \sim$ r.v. exponential with mean 50θ

$$f \sim \frac{1}{50\theta} e^{-\frac{X(\theta)}{50\theta}} \text{ for all methods.}$$

$$Var[X_n(\theta)] = (50\theta)^2 \text{ is bounded function of } \theta$$

$$G(\theta) = -\left(\frac{d}{d\theta} \mathbb{E}[h(X|\theta)) - 1 \right)$$

For $\boxed{IPA}$: in example 7.3 said that $\mathbb{E}\left[\frac{X(\theta)}{\sqrt{1+X(\theta)}}\right]$ is convex and has unique min θ^*

For $\boxed{SF}$:

$$G(\theta) = -\left[\frac{X(\theta)}{(X(\theta)+1)^{\frac{1}{2}}}\frac{1}{\theta}\left(\frac{X(\theta)}{50\theta}-1\right)-1\right]$$

$\mathbb{E}(X_n) = 50\theta$

$Var(X_n(\theta)) = (50\theta)^2$ – bounded in θ

$G(\theta)$ is monotonic on $(0; +\infty]$, $G(\theta)$ is strictly decreasing $\Rightarrow$ we have unique θ^*

$G''(\theta) < 0 \Rightarrow$ So we have unique globaly assymptotic stable point $\lim_{t\to\infty} x(t) = \theta^*$

Also, $G'(\theta) < 0 \Rightarrow$ algorithm is converges to ODE $\frac{dx(t)}{dt} = G(x(t))$

For: $\boxed{MVD}$:

~~$G(\theta) = -\left[\frac{1}{\theta}\mathbb{E}\left(Gamma\left(2, \frac{1}{50\theta}\right)\right) - \frac{1}{\theta}\mathbb{E}\left[\frac{1}{}\right.\right.$~~

$$G(\theta) = -\left[\frac{1}{\theta}\mathbb{E}\left(h(X^+(\theta)) - h(X^-(\theta))\right) - 1\right]$$

$h = \frac{X(\theta)}{\sqrt{1+X(\theta)}}$ $\quad X^+ \sim Gama\left(2, \frac{1}{50\theta}\right)$ $x > 50\theta$

$X^- \sim exp\left(-\frac{X(\theta)}{50\theta}\right)\cdot\frac{1}{50\theta}$ $x \le 50\theta$

$Var(X_n^-(\theta)) = (50\theta)^2$ – bounded in θ on $(x \le 50\theta)$

$Var(X_n^+(\theta)) = 2(50\theta)^2$ bounded ~~on~~ in θ on $(x > 50\theta)$

$G(\theta)$ is coersive $\Rightarrow$ we have unique θ^*

in IPA, SF, and MVD we have

$X(\theta)$ has exponential distribution with mean 50θ

using Theorem 6.2 we have

$$\mathbb{E}[\|\theta_n - \theta^*\|^2] = O(n^{-k}) \quad k = \min(2\beta, \gamma + \delta)$$

$\varepsilon_n = n^{-\gamma}$; $\|B_n\| = O(n^{-\beta})$ $\|V_n\| = O(n^{-\delta})$

no bias in our estimators so $\beta = +\infty$ $\quad B_n = 0$

$\varepsilon_n = n^{-\gamma}$ for $\gamma \in (0, 1]$

so $k = \gamma + \delta$

$$\mathbb{E}(\bar{X}_n) = 50\theta$$

$$\mathrm{Var}(\bar{X}_n) = \frac{1}{n} \mathrm{Var}(\bar{X}_n) = \frac{1}{n}(50\theta)^2 \Rightarrow \delta = 1$$

where $\bar{X}_n$ sample average $\bar{X}_n = \frac{1}{n}\sum_i^n X_i$

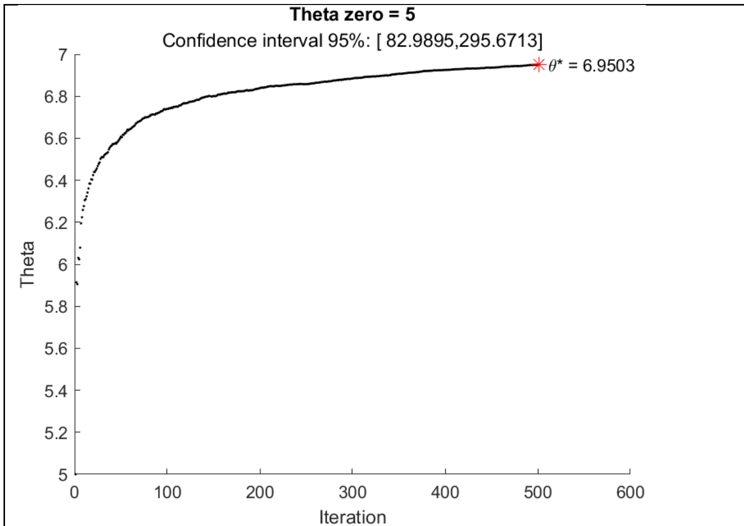
max of k will be when $\gamma = 1$, $k = 2$.

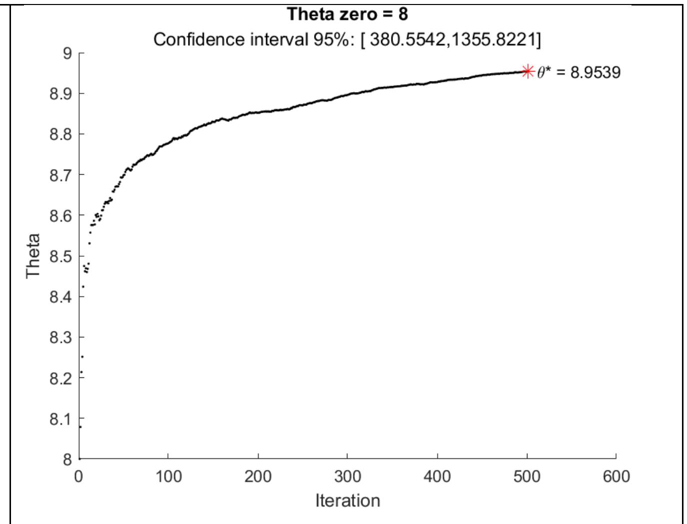
$\varepsilon_n = \frac{1}{n}$

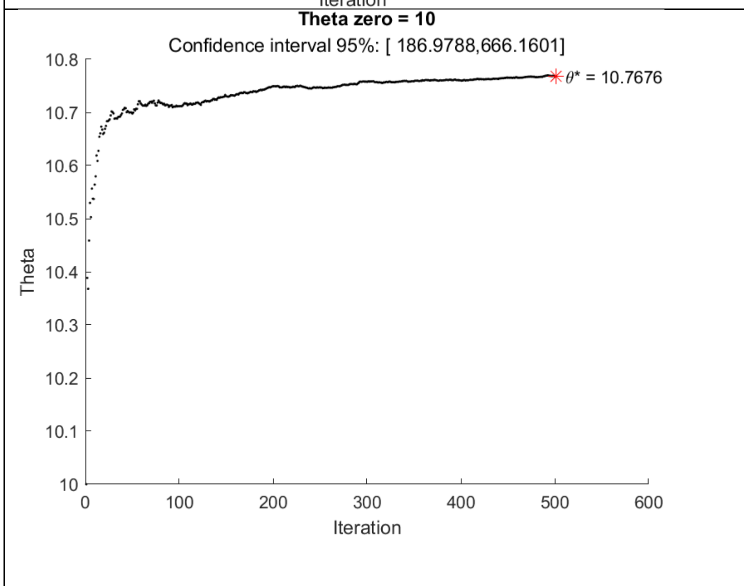
a) using simulation we find out the CPU time for the
IPA, N = 500 iterations for the
theta = [5,8,10,12,15]; we have relevant CPU time in seconds
CPU time =[0.0225 0.0243 0.0268 0.0204 0.0212] around 0.02 seconds. Also we used batch size in 10 sample of random variable with mean 50*theta

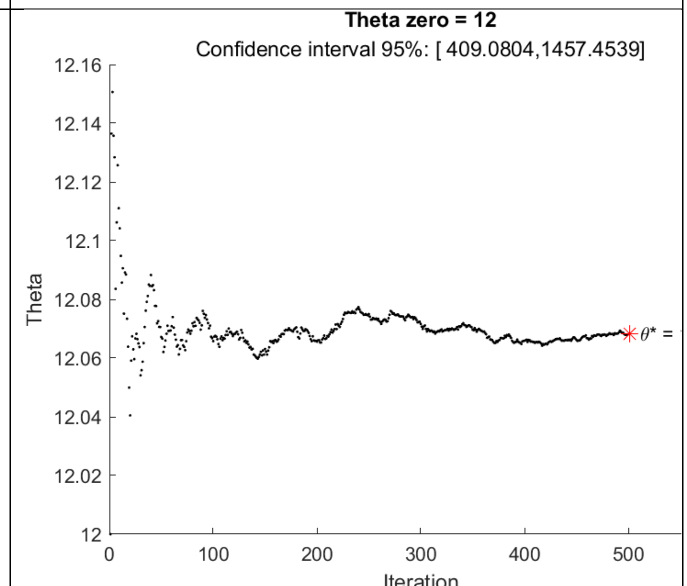
For the SF

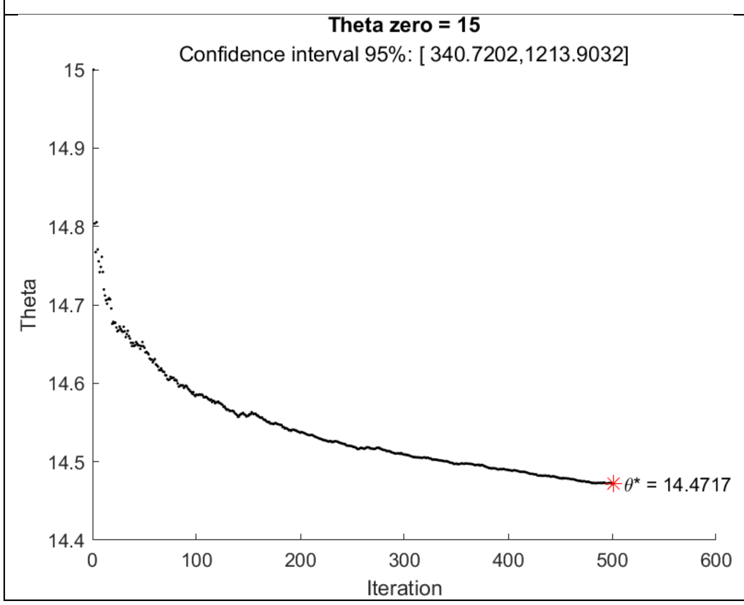
b) following table shows all the graphs for the initial theta give as in exercise.











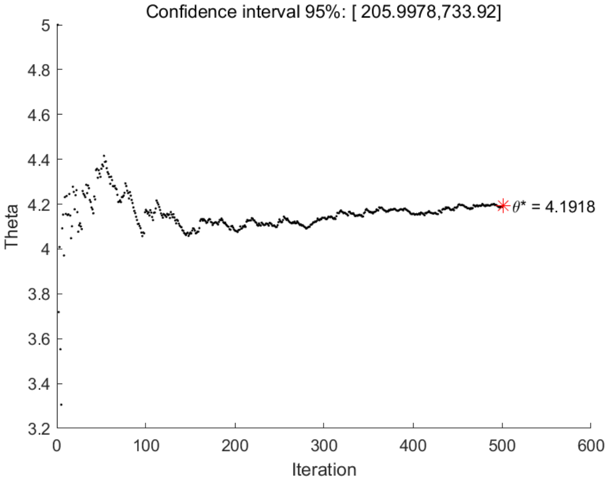
Graphs for the IPA methods, for N =500 iteration , CPU time is around 0.02 seconds

SF method

for N =500 iteration , CPU time is [0.03480 0.02144 0.02 0.01908 0.0227] around 0.02 seconds

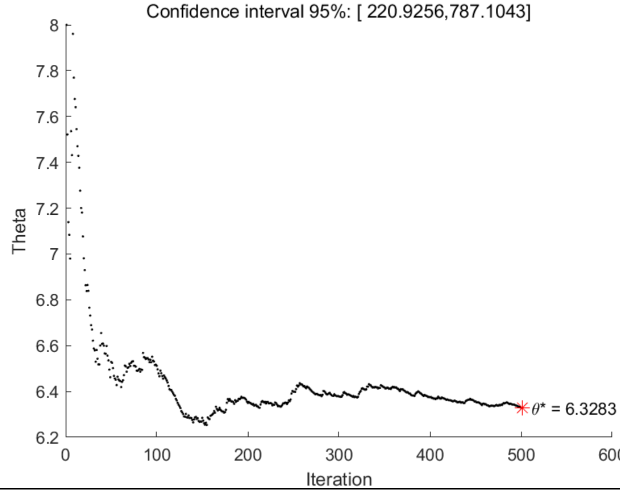
Theta zero = 5

Confidence interval 95%: [205.9978,733.92]



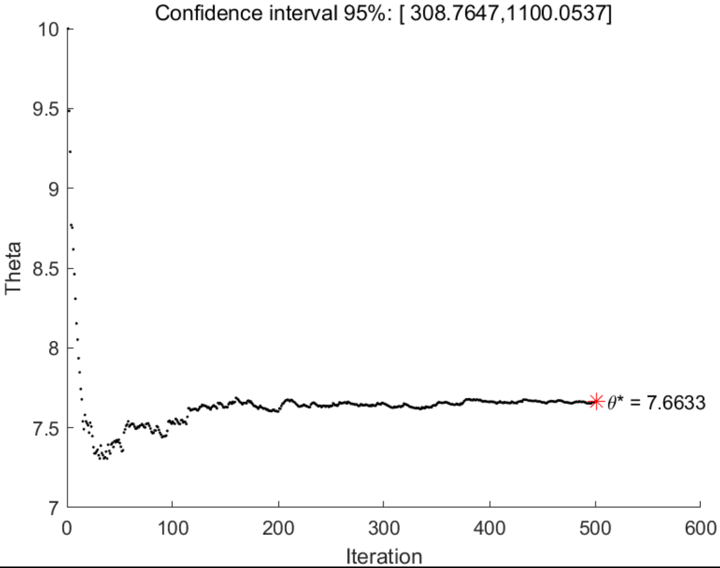
Theta zero = 8

Confidence interval 95%: [220.9256,787.1043]



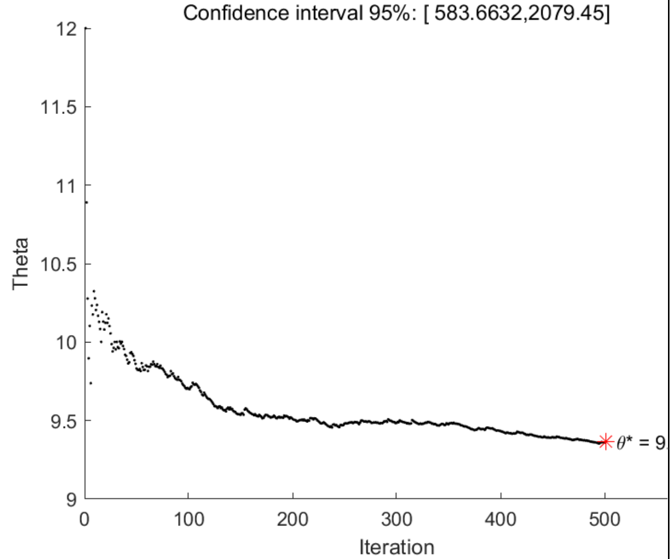
Theta zero = 10

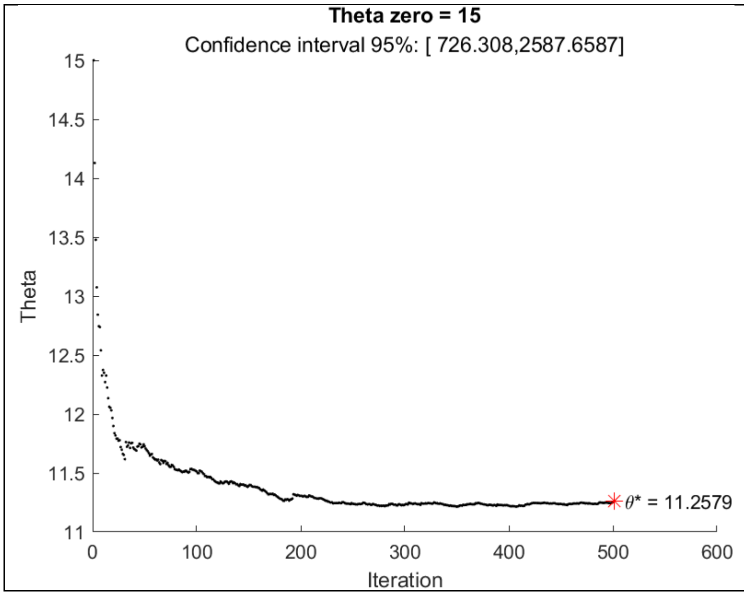
Confidence interval 95%: [308.7647,1100.0537]



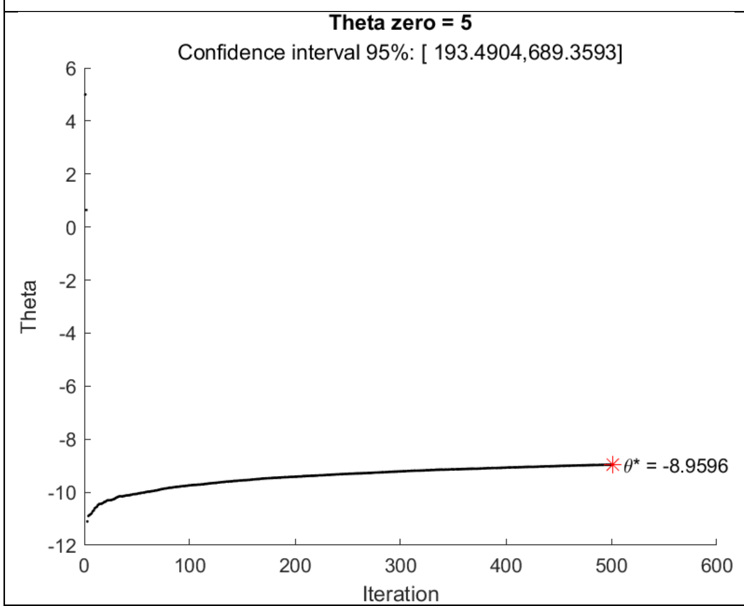
Theta zero = 12

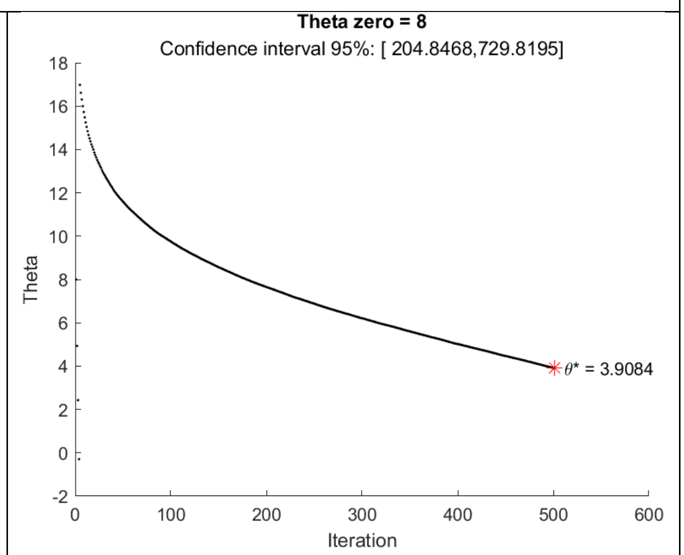
Confidence interval 95%: [583.6632,2079.45]

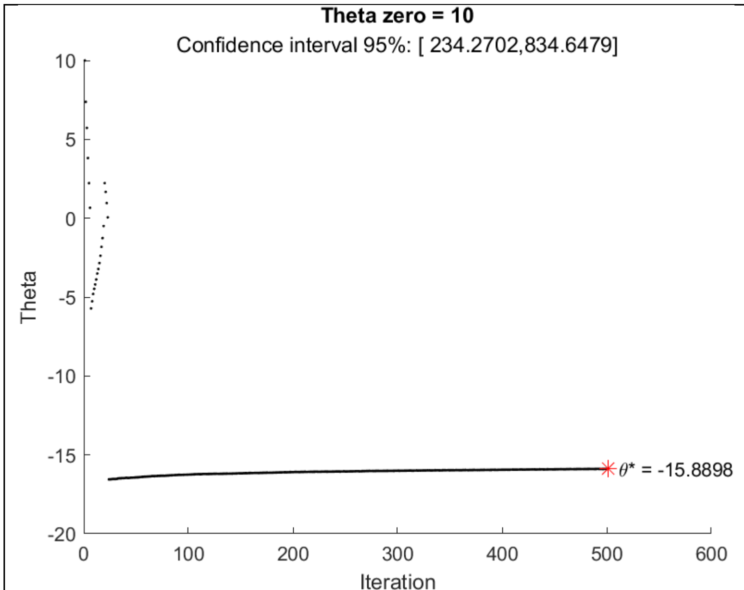


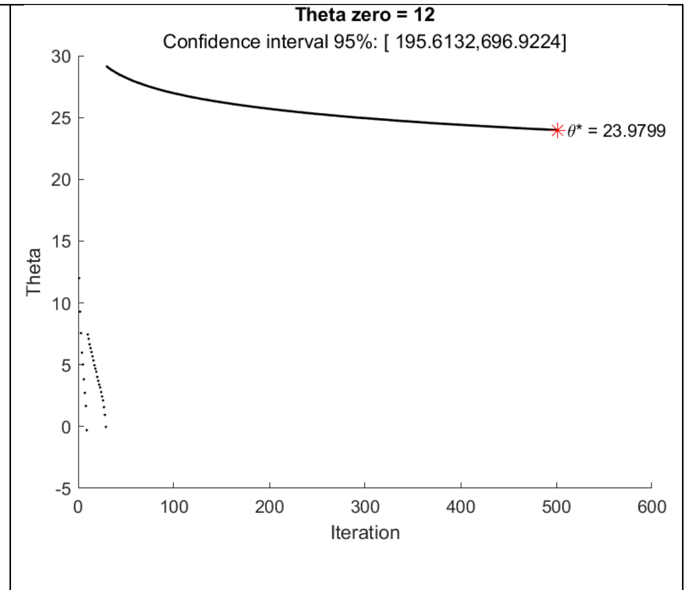

Theta zero = 15
Confidence interval 95%: [726.308,2587.6587]
Theta
Iteration
θ* = 11.2579

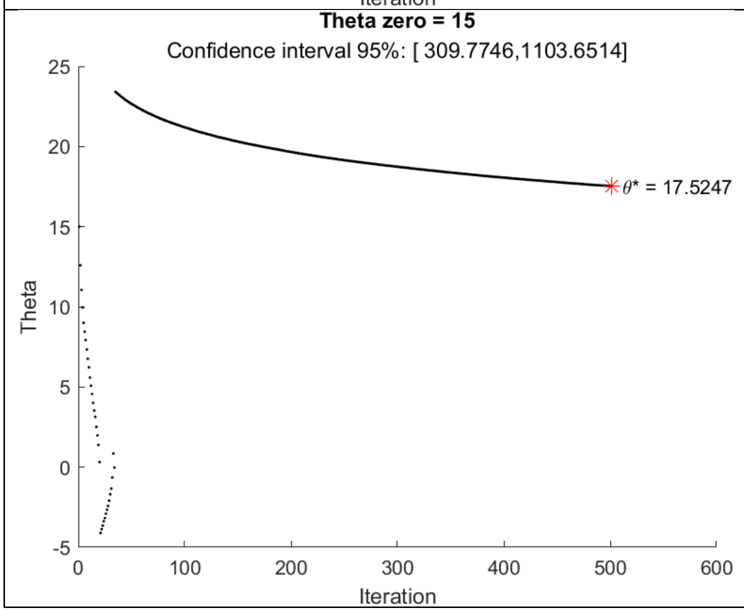
MVD


Theta zero = 5
Confidence interval 95%: [193.4904,689.3593]
Theta
Iteration
θ* = -8.9596


Theta zero = 8
Confidence interval 95%: [204.8468,729.8195]
Theta
Iteration
θ* = 3.9084







This method is not working correct, but we could not find mistake. The cpu run time is also bigger compared to the IPA and SF methods: [1.12192 0.77141 0.7548 0.6736 0.516269] in seconds.
When theta zero is grow the cpu time is smaller that should not be the case, probably algorithm is not correct.

IPA	SF method	MVD
CPU time = 0.02 seconds	CPU time = 0.02 seons	CPU time = around 1 second
For theta zero = 15 Theta = 14.5	For theta zero = 15 Theta = 11.25	For theta zero =15 Theta = 17.5

EXERCISE 8.11. Refer to Exercise 7.7 on the optimal strategic investment for the mining model. Show that the IPA, the SF and the MVD methods provide unbiased estimators for the derivative.

Exercise 8.11.

We apply theorem 8.1 to show that estimators in IPA, SF, MVD methods are unbiased.

Thus we need show that all conditions of T. 8.1 holds.

IPA: Using represatem $X(\theta) \approx 50\theta \ln(1-u)$ from Example 7.2 we have

$$\frac{dX(\theta)}{d\theta} = \frac{X(\theta)}{\theta} \Rightarrow \text{exist (i)}. \checkmark$$

h is differentiable mappin with derivative $\frac{d}{d\theta} h(X(\theta)) = \frac{1}{\theta} X(\theta) h'(X(\theta))$

$X(\theta) = 50\theta \ln(1-u)$ - monotone increasing in θ

so h ~~has~~ monotone increasing derivative

$$\sup_{\theta\in(\theta_l,\theta_r)} \frac{d}{d\theta} h(X(\theta)) \le \frac{1}{\theta_l} X(\theta_r) h'(X(\theta_r)) \le \frac{1}{\theta_l} X(\theta_r) |h'(X(\theta_r))| = K$$

So $E[X(\theta_r)|h'(X(\theta_r))|]$ is finite $\Rightarrow$

$h(X(\theta))$ - is Lipschitz continous (iii).

IPA estimator is unbiased.

SF: ~~$X(\theta) \approx 50\theta \ln(1-u)$~~

~~Representation is exist~~, such that

~~$\frac{dX(\theta)}{d\theta} = \frac{X(\theta)}{\theta}$ so~~ and derivative ~~exist~~. (i)

Cummulative distribution function of Exponential with mean 50θ

$$F(\theta) = 1 - e^{-\frac{X(\theta)}{50\theta}}$$

$F(\theta)$ is continuously differentiable with respect to θ and continously differentiable with respect to X

Then $\frac{d}{d\theta}X(\theta) = -\dfrac{\frac{\partial}{\partial\theta}F_\theta(X(\theta))}{\frac{\partial}{\partial x}F_\theta(X(\theta))} = + \dfrac{+e^{-\frac{X(\theta)}{500}} \cdot \left(+\frac{X(\theta)}{500\theta^2}\right)}{+e^{-\frac{X(\theta)}{500}} \cdot \left(+\frac{1}{500}\right)} =$

$= \dfrac{X(\theta)}{\theta}$ - exist (i)

$h = \dfrac{X(\theta)}{\sqrt{X(\theta)+1}}$ - is differentiable on $[0; +\infty)$. (ii)

$$h' = \frac{1}{(1+X(\theta))^{\frac{1}{2}}} - \frac{X(\theta)}{2(1+X(\theta))^{3/2}}$$

as $X \to \infty$

$$h' \to \underbrace{\frac{1}{(1+X(\theta))^{\frac{1}{2}}}}_{\to 0} - \frac{X(\theta)}{2X^{3/2}\left(\underbrace{\frac{1}{X^{3/2}}}_{\to 0}+1\right)^{3/2}} \Rightarrow$$

$\Rightarrow h' < \infty \Rightarrow$ suffitien condition to have unbiased estimator

so $E(X(\theta_r) \mid h'(X(\theta_r)) < \infty \Rightarrow$ (iii) holds.

$\boxed{\text{MVD}}$. we have $f_\theta = \frac{1}{500}e^{-\frac{X(\theta)}{500}}$ is differentiable on $[0, \infty)$

θ is strictly monotone incrisng.

CDF exis and $\frac{d}{d\theta}X(\theta)$ - exist as shown above.

h' is exis and bounded $\Rightarrow$

$h(X(\theta))$ Lipshitz continuous.

Exercise 8.13. Refer to the reliability model of Examples 7.15, 8.4 and 8.17, illustrated in Figure 8.3. Consider now the model where the lifetimes have a Weibull distribution with parameters (λ_i, k). In particular, for component 3 we have: $\mathbb{P}(T_3 \le x) = G_3(x) = 1 - e^{-(x/\theta)^k}$, so that now $\mathbb{E}(T_3) = \theta\,\Gamma(1 + 1/k)$.

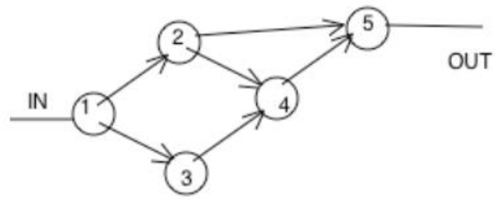


Figure 8.3: Reliability Network

(a) Find the representation $T_3 = G_3^{-1}(U)$ and show that the IPA estimator:

$$\widehat{L}^{\mathrm{IPA}}(\theta,\omega) = \begin{cases} \frac{T_3(\theta,\omega)}{\theta} & T_2(\omega) < T_3(\theta,\omega) < \bar{T}(\omega) = \min(T_1(\omega), T_4(\omega), T_5(\omega)) \\ 0 & \text{otherwise} \end{cases} \tag{8.36}$$

is valid for this case. For what family of distributions G_3 will the IPA estimator remain unchanged?

(b) Explain how the SF and MVD estimators of Exercise 8.12 should be modified for this problem.

(8.13)

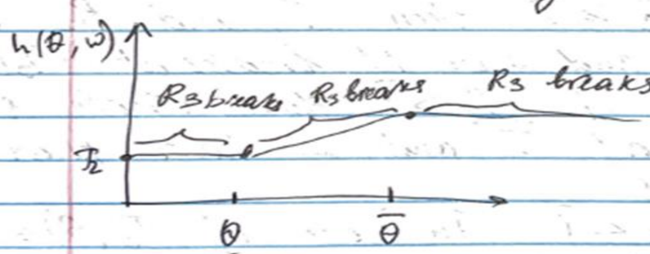
So the system will at route R_1 and R_2 will breaks at T_2 and R_3 will breakes at $T_3(\theta)$. If $T_3(\theta) \geq \overline{T}$ in this's case since $T_2 < \overline{T}$ then R_1 and R_2 will break prior to R_3 breaking at $\overline{T}$ System life time in this's case will be $\overline{T}$.

Case 3: $T_2 < T_3(\theta) < \overline{T}$ we will have R_1 and R_2 to break at T_2, R_3 breaks at $T_3(\theta)$ then the system's life will be:

$T_2 < \overline{T} \Rightarrow \underline{\theta} = -T_2(\omega) / \frac{1}{1 - e^{\frac{(-1)^k x^k}{\omega^k}}}$, $\overline{\theta} = -\overline{T}(\omega) / \frac{1}{1-e^{(-\frac{x}{\omega})^k}}$

then
$$h(\theta, \omega) = \begin{cases} T_2(\omega) & \theta \leq \underline{\theta} \\ \frac{1}{1-e^{\frac{(-1)^k x^k}{\omega^k}}} & \underline{\theta} \leq \theta \leq \overline{\theta} \\ \overline{T}(\omega) & \theta \geq \overline{\theta} \end{cases}$$

This function is continues and then Lipschitz constant is bounded to an absolute value by $\frac{1}{1-e^{-(\frac{x}{\omega})^k}}$



Therefor according to Theorem 8.2, $L^{IPA}(\theta)$ is indenependent of the $T_3(\theta)$ and because $T_3(\theta) = (1 - e^{-(\frac{x}{\theta})^k})^{-1}$ is sufficient to establish the result.

Possible paths : $R_1 = \{1, 2, 5\}$
$R_2 = \{1, 2, 4, 5\}$
$R_3 = \{1, 3, 4, 5\}$

T_3 distribution is $P(T_3 \le x) = G_3(x) = 1 - e^{-(x/\theta)^K}$
so $\mathbb{E}(T_3) = \theta\Gamma(1 + 1/K)$

(a) $T_3 = G_3^{-1}(V)$ where $V = V(0,1)$
since $G_3(x) = 1 - e^{(-x/\theta)^K}$ then
distribution $T_3 = G_3^{-1} = (1 - e^{(-x/\theta)^K})^{-1} = \dfrac{1}{1 - e^{(-x/\theta)^K}}$

$$L^{IPA}(\theta, w) = \begin{cases} \dfrac{T_3(\theta, w)}{\theta}, & T_2(w) < T_3(\theta, w) < \overline{T}(w) = \min(T_1(w), T_4(w), T_5(w)) \\ 0, & \text{otherwise} \end{cases}$$

$$\text{or } L^{IPA}(\theta, w) = \begin{cases} \dfrac{\frac{1}{1 - e^{(-x/\theta)^K}}}{\theta}, & \text{if } T_2(w) < T_3(\theta, w) < \overline{T}(w) \\ & \text{iff } T_2(w) < \frac{1}{1 - e^{(-x/\theta)^K}} < \overline{T}(w) \\ 0, & \text{otherwise} \end{cases}$$

for each L there is a piece-wise linear function of θ

Case 1: $T_2 \ge \overline{T}$ then R_1 and R_2 don't depend on R_3 and when it breaks at $(T_3(\theta), \overline{T})$ if $T_3(\theta) < \overline{T}$. if $T_3 > \overline{T}$ then whole system will break at $\overline{T}$ and we have that $\overline{T}$ depends on θ

Case 2: $T_2 < \overline{T}$ In case if $T_3(\theta) \le T_2$ and $T_3 < \overline{T}$ then R_3 will breaks at earliest time possible a time $T_3(\theta)$ but again R_1 and R_2 don't depend on R_3 and its break

$$L^{IPA}(\theta, w) = \begin{cases} \dfrac{\left(1-e^{-\left(\frac{x}{\theta}\right)^k}\right)^{-1}}{\theta} & T_2(w) \le T_3(\theta, w) < \bar{T}(w) = \min(T_1(w), T_4(w), T_5(w)), \\ 0 & \text{otherwise} \end{cases}$$

IPA is unbiased when θ is a scale parameter of the distribution so we don't need to know distribution of T_3 (according to example 8.17)

(b) Using lemma 8.1 $X(\theta)$ derivative will be denoted by $X'(\theta)$ and can be obtain from $\dfrac{-\frac{\partial}{\partial \theta} F_\theta(X(\theta))}{f_\theta(X(\theta))}$

$$\Rightarrow E'[X'(\theta) \mid X(\theta)] = \frac{-\frac{\partial}{\partial \theta} F_\theta(X(\theta))}{f_\theta(X(\theta))}$$

According to example 8.17.

$$\frac{d}{d\theta} E[L(T)] = \int h(t) \frac{\partial}{\partial \theta}\left(\frac{1}{\theta} \cdot \left(1-e^{-\left(\frac{x}{\theta}\right)^k}\right)^{-1}\right) dt$$

(using product rule)

$$\frac{\partial}{\partial(\theta)}\left(\frac{1}{\theta}\left(1-e^{-\left(\frac{x}{\theta}\right)^k}\right)^{-1}\right) = \frac{d}{d(\theta)}[f(\theta) g(\theta)] =$$

$$= f(\theta) \frac{d}{d\theta}[g(\theta)] + g(\theta) \frac{d}{d\theta}[f(\theta)]$$

$$\Rightarrow -\theta^{-2}\left(1-e^{-\frac{kx}{\theta}}\right)^{-2} \cdot \left(e^{-\frac{kx}{\theta}}(kx(-\theta^{-1})) + 1 - e^{-\frac{kx}{\theta}}\right)$$

which should correspond between $G\left(2, 1-e^{-\left(\frac{x}{\theta}\right)^k}\right)$ and exponential $\left(\frac{1}{\theta}\right)$ distribution.

so corresponding MVD estimator will be:

$$L^{MVD} = \frac{1}{\theta}\left(L\left(T_1, T_2, 1-e^{-\left(\frac{X}{\theta}\right)^K} + X(\theta), T_4, T_5\right) - \right.$$
$$\left. - L\left(T_1, T_2, 1-e^{-\left(\frac{X}{\theta}\right)^K}, T_4, T_5\right)\right)$$

with $X(\theta) \sim \exp(1/\theta)$

SF: from $F_\theta(x) = E[T_3] = \theta\Gamma(1+1/K)$

$$\mathbb{E}[X'(\theta)|X(\theta)] = -\frac{c_\theta}{f_\theta(X(\theta))}\mathbb{E}\left[\theta\Gamma\left(1+\frac{1}{K}\right)SF_\theta\right.$$
$$\left.\tilde{X}(\theta))\right]$$ where $\tilde{X}(\theta)$ is a copy $X(\theta)$

EXERCISE 8.20. Train A arrives at the central station at time $X(\theta)$ and it is next schedule for departure at time s. The delay is $\max(0, X(\theta) - s) = (X(\theta) - s)_+$. The parameter θ governs the distribution of $X(\theta)$ and can in principle be adjusted (speed of train is one example, make and model, or driver's experience are other examples), perhaps satisfying some constraints. In order to carry out an optimization procedure it is first necessary to estimate the *sensitivity* or derivative of the expected delay with respect to θ. In other words, we wish to estimate

$$\frac{d}{d\theta}\mathbb{E}[h(X(\theta))]; \quad h(x) = \max(0, x - s).$$

For this exercise consider $X(\theta) \sim \text{Erlang}(3, 1/\theta)$, with density $f_\theta(x) = (x^2/2\theta^3)e^{-x/\theta}$ for $x > 0$.

(a) Calculate the score function for this distribution.

(b) Verify the conditions of Theorem 8.3. What family of functions $v(\cdot)$ can be used? Is h in the family?

(c) Give the simulation algorithm for generating n replications of $D^{\text{SF}}(\theta)$. From these n replications, give the sample average, the sample variance, and provide the formula for estimating a confidence interval at confidence level α.

a). $f_\theta(x) = \left(\frac{x^2}{2\theta^3}\right) e^{-\frac{x}{\theta}}$

$$SF = \frac{\partial}{\partial\theta} \ln\left(f_\theta(x)\right) = \frac{\partial}{\partial\theta} \ln\left(\frac{x^2}{2\theta^3} e^{-\frac{x}{\theta}}\right) =$$

$$= \frac{\cancel{\frac{x^2}{2}} \frac{(-3)}{\theta^4} \cancel{e^{-\frac{x}{\theta}}} + \cancel{\frac{x^2}{2\theta^3}} \cdot \cancel{e^{-\frac{x}{\theta}}} \cdot \left(+\frac{x}{\theta^2}\right)}{\cancel{\frac{x^2}{2\theta^3}} \cancel{e^{-\frac{x}{\theta}}}} = \left(-\frac{3}{\theta} + \frac{x}{\theta^2}\right) = SF$$

EXERCISE 8.21. Consider $L(\theta) = \mathbb{E}\big[h\big(X(\theta)\big)\big]$, where $h\big(X(\theta)\big) = \max\big(X(\theta) - s, 0\big)$, for some given $s > 0$, and where $X(\theta)$ has an Erlang- $(3, 1/\theta)$-distribution for $\theta > 0$ with pdf

$$\frac{1}{2\theta^3} x^2 e^{-x/\theta}, x \geq 0.$$

(a) Derive the MVD estimator $D^{\mathrm{MVD}}(\theta)$ of $L'(\theta)$.

[*Hint:* it is a difference of two estimators involving Erlang- $(4, 1/\theta)$ and Erlang-$(3, 1/\theta)$ distributions.]

(b) Give the simulation algorithm for generating n replications of $D^{\mathrm{MVD}}(\theta)$, including how you generate from the Erlang-$(4, 1/\theta)$ and Erlang-$(3, 1/\theta)$ distributions. Make sure to exploit common random numbers as much as possible. From these n replications, give the sample average, the sample variance, and provide the formula for estimating a confidence interval at confidence level α.

(c) Give the expression for a randomized MVD estimator $D^{\mathrm{MVDrand}}(\theta)$ that involves a single estimator (instead of the difference in (b)). Show that $\mathbb{E}\big[D^{\mathrm{MVDrand}}(\theta)\big] = \mathbb{E}\big[D^{\mathrm{MVD}}(\theta)\big]$.

Excersize 8.21

$L(\theta) = E[h(X(\theta))]$

$h(X(\theta)) = \max(X(\theta) - s, 0)$ $\quad s>0$

$f_\theta(x) = \frac{1}{2\theta^3} x^2 e^{-x/\theta}$ $\quad x \geq 0$

a) Find MVD Estimator $D^{MVD}(\theta)$ of $L'(\theta)$

$$L'(\theta) = \frac{d}{d\theta} E[h(X(\theta))] = \int h(X(\theta)) \frac{d}{d\theta} f_\theta(x)\, dx$$

$$\frac{d}{d\theta} f_\theta(x) = \frac{d}{d\theta} \frac{1}{2\theta^3} x^2 e^{-x/\theta}$$

take out constants $\frac{1}{2}$ and x^2

$$= \frac{d}{d\theta} xy = \frac{d}{d\theta} x \cdot y + x \cdot \frac{d}{d\theta} y$$

where $x = \frac{1}{\theta^3}$ $\quad y = e^{-x/\theta}$

$$\frac{d}{d\theta} \frac{1}{\theta^3} = -\frac{3}{\theta^4}$$

take out constant $-x$

$$\frac{d}{d\theta} e^{-x/\theta} = e^{-x/\theta} \frac{d}{d\theta} - \frac{x}{\theta} \qquad \frac{d}{d\theta} \frac{1}{\theta} = -\frac{1}{\theta^2}$$

bring back $-x$, $\frac{x}{\theta^2}$

$$= e^{-x/\theta} \frac{x}{\theta^2}$$

bring back $\frac{1}{2}$ and x^2

$$\frac{d}{d\theta} f_\theta(x) = \frac{x^2}{2}\left(-\frac{3}{\theta^4} e^{-x/\theta} + \frac{1}{\theta^3} e^{-x/\theta} \frac{x}{\theta^2}\right)$$

$$= \frac{x^2}{2} \cdot -\frac{3}{\theta^4} \cdot e^{-x/\theta} + \frac{x^2}{2} \cdot \frac{1}{\theta^3} \cdot e^{-x/\theta} \cdot \frac{x}{\theta^2}$$

$$= -\frac{3x^2 e^{-x/\theta}}{2\theta^4} + \frac{x^3 e^{-x/\theta}}{2\theta^5}$$

$$= \frac{x^2 e^{-x/\theta}}{2\theta^5} [x - 3\theta]$$

$$f_\theta^+(x) = \frac{x^2 e^{-x/\theta}}{2\theta^5}(x-3\theta) \quad \text{when } x > 3\theta$$

$$f_\theta^-(x) = \frac{x^2 e^{-x/\theta}}{2\theta^5}(3\theta-x) \quad \text{when } 3\theta > x$$

$$C_\theta = \int_{3\theta}^{\infty} f_\theta^+(x) \quad \text{or} \quad \int_{3\theta}^{\infty} f_\theta^-(x)$$

they are equal because a pdf always integrates to 1 and both parts cancel out, so changing one changes the other. And a PDF is always non negative so for $f_\theta^-(x)$ we take the opposite of the negative.

$$C_\theta = \int_{3\theta}^{\infty} \frac{x^2 e^{-x/\theta}}{2\theta^5}(x-3\theta) \quad \text{take out constants } \frac{1}{2\theta^5}$$

$$= \int_{3\theta}^{\infty} x^3 e^{-x/\theta} - \int_{3\theta}^{\infty} 3\theta x^2 e^{-x/\theta}$$

$$\int x^m e^{ax}\,dx = e^{ax} \sum_{r=0}^{m} (-1)^r \frac{m!\, x^{m-r}}{(m-r)!\, a^{r+1}}$$

evaluate this at $x = \infty$ minus at $x = 3\theta$

$$\int_{3\theta}^{\infty} x^3 e^{-x/\theta} \qquad m = 3 \quad a = -\frac{1}{\theta}$$

$$x \text{ at } \infty : e^{-\infty/\theta} \sum_{r=0}^{3} (-1)^r \frac{3!\, \infty^{3-r}}{(3-r)!\, -\frac{1}{\theta}^{r+1}}$$

As we near infinity e^{-x} goes to 0 faster than x^n goes to infinity so the product goes to 0.

$$X \text{ at } 3\theta : e^{-3\theta/\theta} \sum_{r=0}^{3} (-1)^r \frac{3!\,3\theta^{3-r}}{(3-r)!\,-\frac{1}{\theta}^{r+1}}$$

$$= e^{-3}\left(-\frac{18\theta^3}{6\frac{1}{\theta}} - \frac{18\theta^2}{2\frac{1}{\theta^2}} - \frac{18\theta}{\frac{1}{\theta^3}} - \frac{18}{\frac{1}{\theta^4}}\right)$$

$\cdot\theta$ $\cdot 3\theta^2$ $\cdot 6\theta^3$ $\cdot 6\theta^4$

$$= e^{-3}\left(-\frac{18\theta^4}{6} - \frac{54\theta^4}{6} - \frac{108\theta^4}{6} - \frac{108\theta^4}{6}\right)$$

$$= e^{-3}\left(\frac{288\theta^4}{6}\right) = e^{-3}(-48\theta^4)$$

$$\int_{3\theta}^{\infty} x^3 e^{-x/\theta} = 0 - e^{-3}(-48\theta^4) = -e^{-3}(-48\theta^4)$$

$$\int_{3\theta}^{\infty} 3\theta x^2 e^{-x/\theta} \quad m=3 \quad a=-\frac{1}{\theta}$$ take out constant 3θ

X at ∞ : goes to 0 like in the other integral

$$X \text{ at } 3\theta : e^{-3\theta/\theta} \sum_{r=0}^{2} (-1)^r \frac{2!\,3\theta^{2-r}}{(2-r)!\,-\frac{1}{\theta}^{r+1}}$$

$$= e^{-3}\left(-\frac{6\theta^2}{2\frac{1}{\theta}} - \frac{6\theta}{\frac{1}{\theta^2}} - \frac{6}{\frac{1}{\theta^3}}\right)$$

$\cdot\theta$ $\cdot 2\theta^2$ $\cdot 2\theta^3$

$$= e^{-3}\left(-\frac{6\theta^3}{2} - \frac{12\theta^3}{2} - \frac{12\theta^3}{2}\right)$$

$$= e^{-3}\left(\frac{-30\theta^3}{2}\right) = e^{-3}(-15\theta^3)$$ bring back 3θ

$$\int_{3\theta}^{\infty} 3\theta x^2 e^{-x/\theta} = 0 - 3\theta e^{-3}(-15\theta^3) = e^{-3}(45\theta^4)$$

$$C_\theta = -e^{-3}(-48\theta^4) - e^{-3}(45\theta^4)$$
$$= -e^{-3}(-48\theta^4 + 45\theta^4) = -e^{-3}(-3\theta^4)$$

$-e^{-3}(-3\theta^4)$ bring back $\frac{1}{2\theta^5}$

$$C_\theta = -e^{-3}\left(-\frac{3\theta^4}{2\theta^5}\right) = e^{-3}\left(\frac{3}{2\theta}\right)$$

$$f+ = \frac{\frac{x^2 e^{-x/\theta}}{2\theta^5}(x-3\theta) \cdot 2\theta}{e^{-3}\left(\frac{3}{2\theta}\right) \cdot 2\theta}$$

divide $\frac{f+}{c\theta}$

$$= \frac{\frac{x^2 e^{-x/\theta}}{\theta^4}(x-3\theta) \cdot e^3}{3e^{-3} \cdot e^3}$$

$$= \frac{\frac{x^2 e^{3-x/\theta}(x-3\theta)}{\theta^4} \cdot \frac{1}{3}}{3 \cdot \frac{1}{3}}$$

$$= \frac{x^2 e^{3-x/\theta}(x-3\theta)}{3\theta^4}$$

$$f- = \frac{x^2 e^{3-x/\theta}(3\theta-x)}{3\theta^4}$$

Since $f+$ and $f-$ are the same except the order of $x-3\theta$ and $3\theta-x$

Now we have c_θ, $f+$, and $f-$!

F+ and f-: Divide each by c theta so they integrate to 1.

B) https://colab.research.google.com/drive/1pS-TyA-9EArxejGjOTiRKA-fjTg8g1Lr?usp=sharing

```python
import random
import math
from statistics import mean, stdev
```

```python
h_thetax = []
N = 1000
S = 2

#theta = 1 so ignoring it
def integral_f_neg(x):
    return 0.5*(x**3) * math.exp(-x)

def integral_f_pos(x):
    return -0.5*(x**3) * math.exp(-x)

def h_x(y):
    return (max(y-S, 0))

negative_space = integral_f_neg(3) - integral_f_neg(0)
positive_space = 0 - integral_f_pos(3)

dict_neg_y = {}
y = 0
while y <= 3:
    dict_neg_y[y] = integral_f_neg(y)/negative_space
    y += .01

y=3
dict_pos_y = {}
while y <=20:
    dict_pos_y[y] = (integral_f_pos(y)-integral_f_pos(3))/positive_space
    y += .01
dict_pos_y[20] = 1
```

```python
int_1 = 0
for n in range(1,N):
    z = random.uniform(0,1)
    u = random.uniform(0,1)
    if u > 0.5:
        #positive space
        for key in dict_pos_y:
            if dict_pos_y[key] <= z:
                int_1 = key
        int_2 = list(dict_pos_y)[list(dict_pos_y).index(int_1) + 1]
        y = (int_1+int_2)/2
        h_thetax.append(h_x(y))
    else:
        #negative space
        for key in dict_neg_y:
            if dict_neg_y[key] <= z:
                int_1 = key
        int_2 = list(dict_neg_y)[list(dict_neg_y).index(int_1) + 1]
        y = (int_1+int_2)/2
        h_thetax.append(h_x(y))

#avergae and variance
avg = mean(h_thetax)
var = stdev(h_thetax)**2

#confidence interval
a = avg + stdev(h_thetax)*2
b = avg - stdev(h_thetax)*2
```

Avg: 1.9793143143142828
Var: 5.322885125556232
Confidence Interval: (6.593590043204455, -2.6349614145758897)