UiO : Matematisk institutt

Det matematisk-naturvitenskapelige fakultet

STK-4051/9051 Computational Statistics Spring 2024 Exercise 1

Instructor: Odd Kolbjørnsen, oddkol@math.uio.no

2.1 The following data are an i.i.d. sample from a Cauchy(θ, 1) distribution: 1.77, −0.23, 2.76, 3.80, 3.47, 56.75, −1.34, 4.24, −2.44, 3.29, 3.71, −2.40, 4.53, −0.07, −1.05, −13.87, −2.53, −1.75, 0.27, 43.21.

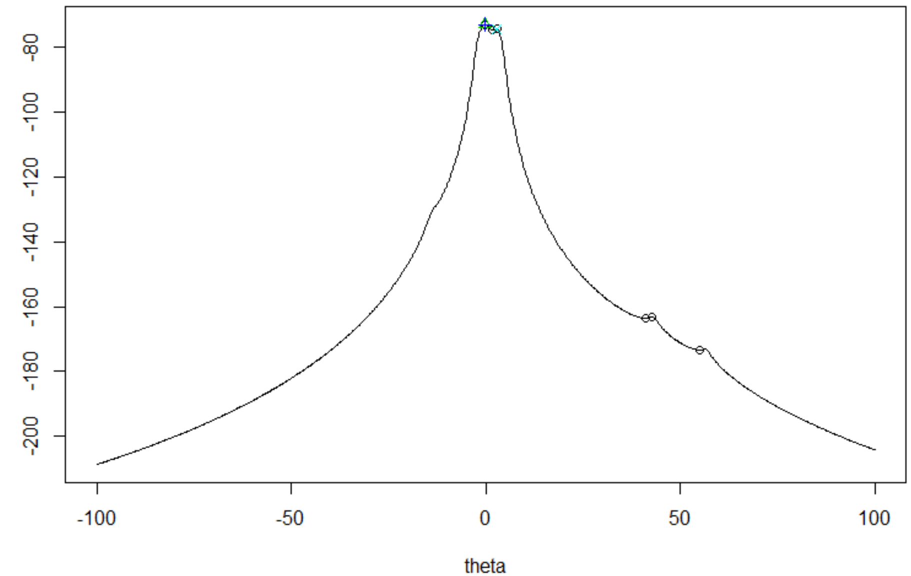
a. Graph the log likelihood function. Find the MLE for θ using the Newton–Raphson method. Try all of the following starting points: −11, −1, 0, 1.5, 4, 4.7, 7, 8, and 38. Discuss your results. Is the mean of the data a good starting point?

b. Apply the bisection method with starting points −1 and 1. Use additional runs to illustrate manners in which the bisection method may fail to find the global maximum.

c. Apply fixed-point iterations as in (2.29), starting from −1, with scaling choices of $\alpha = 1, 0.64$, and 0.25. Investigate other choices of starting values and scaling factors.

d. From starting values of $(\theta^{(0)}, \theta^{(1)}) = (-2, -1)$, apply the secant method to estimate θ. What happens when $(\theta^{(0)}, \theta^{(1)}) = (-3, 3)$, and for other starting choices?

e. Use this example to compare the speed and stability of the Newton–Raphson method, bisection, fixed-point iteration, and the secant method. Do your conclusions change when you apply the methods to a random sample of size 20 from a $N(\theta, 1)$ distribution?



Solution to exercise (2.1)

The Cauchy$(\theta, 1)$ distribution is given by

$$f(x;\theta) = \frac{1}{\pi[1 + (x-\theta)^2]}$$

(a). We have

$$\ell(\theta) = \sum_{i=1}^{n} [-\log(\pi) - \log(1 + (x_i - \theta)^2)]$$

$$\ell'(\theta) = \sum_{i=1}^{n} \frac{2(x_i - \theta)}{1 + (x_i - \theta)^2}$$

$$\ell''(\theta) = -2\sum_{i=1}^{n} \frac{1 - (x_i - \theta)^2}{[1 + (x_i - \theta)^2]^2}$$

$$f(x) = g(x)/h(x)$$

$$f'(x) = \frac{g'(x)h(x) - g(x)h'(x)}{[h(x)]^2}$$

2.3 Let the survival time t for individuals in a population have density function f and cumulative distribution function F. The *survivor function* is then $S(t) = 1 - F(t)$. The *hazard function* is $h(t) = f(t)/(1 - F(t))$, which measures the instantaneous risk of dying at time t given survival to time t. A proportional hazards model posits that the hazard function depends on both time and a vector of covariates, $\mathbf{x}$, through the model

$$h(t|x) = \lambda(t)\exp\left\{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}\right\},$$

where $\boldsymbol{\beta}$ is a parameter vector.

If $\Lambda(t) = \int_{-\infty}^{t} \lambda(u)\, du$, it is easy to show that $S(t) = \exp\left\{-\Lambda(t)\exp\{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}\}\right\}$ and $f(t) = \lambda(t)\exp\left\{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} - \Lambda(t)\exp\{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}\}\right\}$.

a. Suppose that our data are censored survival times t_i for $i = 1, \ldots, n$. At the end of the study a patient is either dead (known survival time) or still alive (censored time; known to survive at least to the end of the study). Define w_i to be 1 if t_i is an uncensored time and 0 if t_i is a censored time. Prove that the log likelihood takes the form

$$\sum_{i=1}^{n}(w_i \log\{\mu_i\} - \mu_i) + \sum_{i=1}^{n} w_i \log\left\{\frac{\lambda(t_i)}{\Lambda(t_i)}\right\},$$

where $\mu_i = \Lambda(t_i)\exp\{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}$.

Consider one censored observation

$$S(t) = \exp\left\{-\Lambda(t)\exp\{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}\}\right\}$$

where $\mu_i = \Lambda(t_i)\exp\{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}$.

If observation i is censored, we have $w_i = 0$ and

$$\begin{aligned} L_i &= \Pr(T \geq t_i) = S(t_i) \\ &= \exp\{-\Lambda(t)\exp(\boldsymbol{x}_i^T\boldsymbol{\beta})\} \\ l_i &= -\Lambda(t)\exp(\boldsymbol{x}_i^T\boldsymbol{\beta}) = -\mu_i \end{aligned}$$

Consider one uncensored observation

$$f(t) = \lambda(t) \exp\left\{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} - \Lambda(t)\exp\{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}\}\right\}.$$

where $\mu_i = \Lambda(t_i)\exp\{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}$.

(a). If observation i is not censored, we have $w_i = 1$ and

$$\begin{aligned}
l_i &= \log(f(t_i)) = \log(\lambda(t_i)) + \boldsymbol{x}_i^T\boldsymbol{\beta} - \Lambda(t_i)\exp(\boldsymbol{x}_i^T\boldsymbol{\beta}) \\
&= \log\left(\frac{\lambda(t_i)}{\Lambda(t_i)}\right) + \log(\Lambda(t_i)) + \boldsymbol{x}_i^T\boldsymbol{\beta} - \mu_i \\
&= \log\left(\frac{\lambda(t_i)}{\Lambda(t_i)}\right) + \log(\Lambda(t_i)\exp(\boldsymbol{x}_i^T\boldsymbol{\beta})) - \mu_i \\
&= \log\left(\frac{\lambda(t_i)}{\Lambda(t_i)}\right) + \log(\mu_i) - \mu_i
\end{aligned}$$

- Likelihood:

$$L\big(\beta,\lambda(t)\big) = \prod_{i=1}^{N} f(t)^{w_i} \cdot S(t)^{1-w_i}$$

$$l\big(\beta,\lambda(t)\big) = \sum_{i=1}^{N} w_i \log\big(f(t)\big) + (1-w_i)\cdot \log(S(t))$$

Insert to get the result

$$\sum_{i=1}^{n} (w_i \log\{\mu_i\} - \mu_i) + \sum_{i=1}^{n} w_i \log\left\{\frac{\lambda(t_i)}{\Lambda(t_i)}\right\},$$

b. Consider a model for the length of remission for acute leukemia patients in a clinical trial. Patients were either treated with 6-mercaptopurine (6-MP) or a placebo [202]. One year after the start of the study, the length (weeks) of the remission period for each patient was recorded (see Table 2.2). Some outcomes were censored because remission extended beyond the study period. The goal is to determine whether the treatment lengthened time spent in remission. Suppose we set $\Lambda(t) = t^\alpha$ for $\alpha > 0$, yielding a hazard function proportional to $\alpha t^{\alpha-1}$ and a Weibull density: $f(t) = \alpha t^{\alpha-1} \exp\left\{\mathbf{x}^T\boldsymbol{\beta} - t^\alpha \exp\{\mathbf{x}^T\boldsymbol{\beta}\}\right\}$. Adopt the covariate parameterization given by $\mathbf{x}_i^T\boldsymbol{\beta} = \beta_0 + \delta_i\beta_1$ where δ_i is 1 if the ith patient was in the treatment group and 0 otherwise. Code a Newton–Raphson algorithm and find the MLEs of α, β_0, and β_1.

c. Use any prepackaged Newton–Raphson or quasi-Newton routine to solve for the same MLEs.

d. Estimate standard errors for your MLEs. Are any of your MLEs highly correlated? Report the pairwise correlations.

e. Use nonlinear Gauss–Seidel iteration to find the MLEs. Comment on the implementation ease of this method compared to the multivariate Newton–Raphson method.

f. Use the discrete Newton method to find the MLEs. Comment on the stability of this method.

where $\mu_i = \Lambda(t_i)\exp\{\mathbf{x}_i^T\boldsymbol{\beta}\}$.

(b). We have in this case

$$\frac{\partial}{\partial\alpha}\Lambda(t) = \frac{\partial}{\partial\alpha}t^\alpha = \frac{\partial}{\partial\alpha}e^{\alpha\log(t)} = e^{\alpha\log(t)}\log(t) = \Lambda(t)\log(t)$$
$$\frac{\partial}{\partial\alpha}\mu_i = \Lambda(t_i)\log(t_i)\exp(\boldsymbol{x}_i^T\boldsymbol{\beta}) = \mu_i\log(t_i)$$
$$\frac{\partial}{\partial\boldsymbol{\beta}}\mu_i = \Lambda(t_i)\exp(\boldsymbol{x}_i^T\boldsymbol{\beta})\boldsymbol{x}_i^T = \mu_i\boldsymbol{x}_i^T$$

$\Lambda(t) = \int_{-\infty}^{t}\lambda(u)\,du,$

$\Lambda(t) = t^\alpha$

$$\log\left(\frac{\lambda(t)}{\Lambda(t)}\right) = \log\left(\frac{\alpha t^{\alpha-1}}{t^\alpha}\right) = \log\left(\frac{\alpha}{t}\right) = \log\alpha - \log t$$

$$\sum_{i=1}^{n}(w_i \log\{\mu_i\} - \mu_i) + \sum_{i=1}^{n} w_i \log\left\{\frac{\lambda(t_i)}{\Lambda(t_i)}\right\},$$

$$\frac{\partial}{\partial \alpha}\Lambda(t) = \frac{\partial}{\partial \alpha}t^{\alpha} = \frac{\partial}{\partial \alpha}e^{\alpha \log(t)} = e^{\alpha \log(t)}\log(t) = \Lambda(t)\log(t)$$

$$\frac{\partial}{\partial \alpha}\mu_i = \Lambda(t_i)\log(t_i)\exp(\boldsymbol{x}_i^T\boldsymbol{\beta}) = \mu_i \log(t_i)$$

$$\frac{\partial}{\partial \boldsymbol{\beta}}\mu_i = \Lambda(t_i)\exp(\boldsymbol{x}_i^T\boldsymbol{\beta})\boldsymbol{x}_i^T = \mu_i \boldsymbol{x}_i^T$$

$$\frac{\partial}{\partial \alpha}l(\boldsymbol{\theta}) = \sum_{i=1}^{n}[(w_i - \mu_i)\log(t_i) + w_i/\alpha]$$

$$\frac{\partial}{\partial \boldsymbol{\beta}}l(\boldsymbol{\theta}) = \sum_{i=1}^{n}(w_i \frac{\mu_i \boldsymbol{x}_i}{\mu_i} - \mu_i \boldsymbol{x}_i) = \sum_{i=1}^{n}(w_i - \mu_i)\boldsymbol{x}_i$$

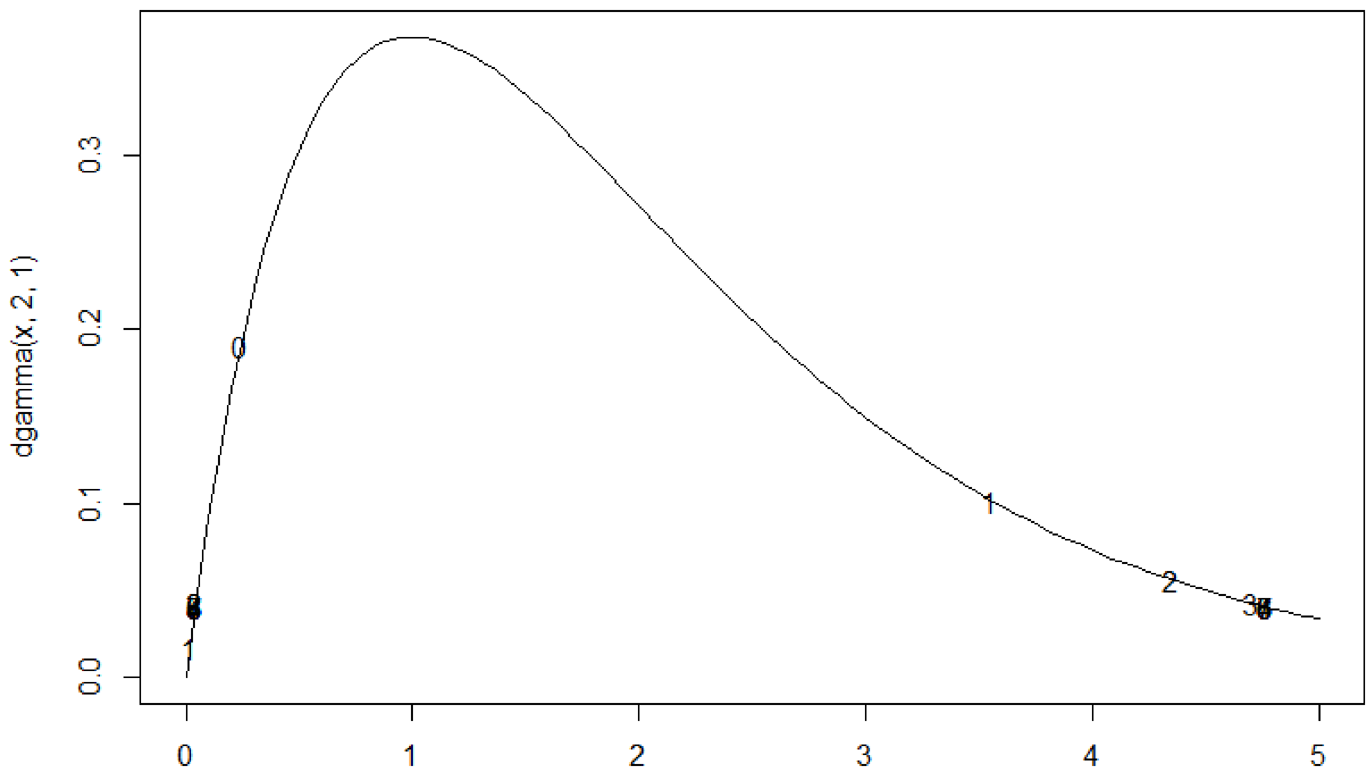
$$\frac{\partial^2}{\partial \alpha^2}l(\boldsymbol{\theta}) = -\sum_{i=1}^{n}[\mu_i \log(t_i)^2 + w_i/\alpha^2]$$

$$\frac{\partial^2}{\partial \boldsymbol{\beta}\partial \boldsymbol{\alpha}}l(\boldsymbol{\theta}) = -\sum_{i=1}^{n}\mu_i \boldsymbol{x}_i \log(t_i)$$

$$\frac{\partial^2}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T}l(\boldsymbol{\theta}) = -\sum_{i=1}^{n}\mu_i \boldsymbol{x}_i \boldsymbol{x}_i^T$$

2.4

A parameter θ has a Gamma(2, 1) posterior distribution. Find the 95% highest posterior density interval for θ, that is, the interval containing 95% of the posterior probability for which the posterior density for every point contained in the interval is never lower than the density for every point outside the interval. Since the gamma density is unimodal, the interval is also the narrowest possible interval containing 95% of the posterior probability.



Solution to exercise (2.4)
We need to find two points x_1, x_2 such that

$$s_1(x_1,x_2) = f(x_2) - f(x_1) = 0$$
$$s_2(x_1,x_2) = F(x_2) - F(x_1) - 0.95 = 0$$

with

$$f(x) = xe^{-x}$$
$$F(x) = \int_0^x f(u)du$$

Now

$$\frac{\partial}{\partial x_1} s_1(x_1,x_2) = -f'(x_1)$$
$$\frac{\partial}{\partial x_2} s_1(x_1,x_2) = f'(x_2)$$
$$\frac{\partial}{\partial x_1} s_2(x_1,x_2) = -f(x_1)$$
$$\frac{\partial}{\partial x_2} s_2(x_1,x_2) = f(x_2)$$

were

$$f'(x) = e^{-x}[1-x]$$

Exercise 1 (Convergence of Newton's method)
(Adapted from (Lange, 2010, exercise 5.13))
Assume we want to extract the root of the function

$$g(x) = x^m - c$$

for $m > 1, c > 0$.

(*a*). Find the (only) exact root of $g(x)$.

(*b*). Show that Newton's method is given by

$$x_n = x_{n-1}\left(1 - \frac{1}{m} + \frac{c}{mx_{n-1}^m}\right).$$

(*c*). Show that $x_n \geq c^{1/m}$ for all $x_{n-1} > 0$.

Hint: This part can be somewhat difficult. Perhaps easiest to start backwards, what requirements are needed for $x\left(1 - \frac{1}{m} + \frac{c}{mx^m}\right) \geq c^{1/m}$ and then relate this to $y = c/x^m$.

(*d*). Show that $x_n \leq x_{n-1}$ whenever $x_{n-1} \geq c^{1/m}$

(*e*). What does this imply if we start with $x_0 > 0$?

Exercise 1 (Convergence of Newton's method)
(Adapted from (Lange, 2010, exercise 5.13))
Assume we want to extract the root of the function

$$g(x) = x^m - c$$

for $m > 1, c > 0$.

(a). Find the (only) exact root of $g(x)$.

(b). Show that Newton's method is given by

$$x_n = x_{n-1}\left(1 - \frac{1}{m} + \frac{c}{m x_{n-1}^m}\right).$$

(a). We find directly that $x^* = c^{1/m}$

(b). We have $g'(x) = m x^{m-1}$ giving

$$\begin{aligned} x_n &= x_{n-1} - \frac{g(x_{n-1})}{g'(x_{n-1})} = x_{n-1} - \frac{x_{n-1}^m - c}{m x_{n-1}^{m-1}} \\ &= x_{n-1} - \frac{x_{n-1}}{m} + \frac{c x_{n-1}}{m x_{n-1}^m} = x_{n-1}[1 - \frac{1}{m} + \frac{c}{m x_{n-1}^m}] \end{aligned}$$

(c). Show that $x_n \geq c^{1/m}$ for all $x_{n-1} > 0$.

Hint: This part can be somewhat difficult. Perhaps easiest to start backwards, what requirements are needed for $x\left(1 - \frac{1}{m} + \frac{c}{mx^m}\right) \geq c^{1/m}$ and then relate this to $y = c/x^m$.

$$x[1 - \frac{1}{m} + \frac{c}{mx^m}] \geq c^{1/m}$$

$$\Updownarrow$$

$$[1 - \frac{1}{m} + \frac{c}{mx^m}] \geq \frac{c^{1/m}}{x} = \left(\frac{c}{x^m}\right)^{1/m}$$

$$\Updownarrow$$

$$1 - \frac{1}{m} + \frac{y}{m} \geq y^{1/m} \qquad y = \frac{c}{x^m}$$

$$\Updownarrow$$

$$1 - \frac{1}{m} \geq y^{1/m} - \frac{y}{m}$$

Denote $h(y) = y^{1/m} - \frac{y}{m}$. Then

$$h'(y) = \frac{1}{m} y^{-\frac{m-1}{m}} - \frac{1}{m}$$

$$h''(x) = -\frac{1}{m}\frac{m-1}{m} y^{-\frac{2m-1}{m}} < 0 \text{ for } y > 0$$

Therefore $h(y)$ has a max point when $h'(y) = 0$ corresponding to $y = 1$ in which case $h(y) = 1 - \frac{1}{m}$. This shows that $1 - \frac{1}{m} \geq y^{1/m} - \frac{y}{m}$ which then implies that $x > c^{1/m}$

(d). Show that $x_n \leq x_{n-1}$ whenever $x_{n-1} \geq c^{1/m}$

(e). What does this imply if we start with $x_0 > 0$?

(d). We have

$$\begin{aligned} x_n =& x_{n-1}[1 - \frac{1}{m} + \frac{c}{m x_{n-1}^m}] \\ \leq& x_{n-1}[1 - \frac{1}{m} + \frac{c}{mc}] = x_{n-1} \end{aligned}$$

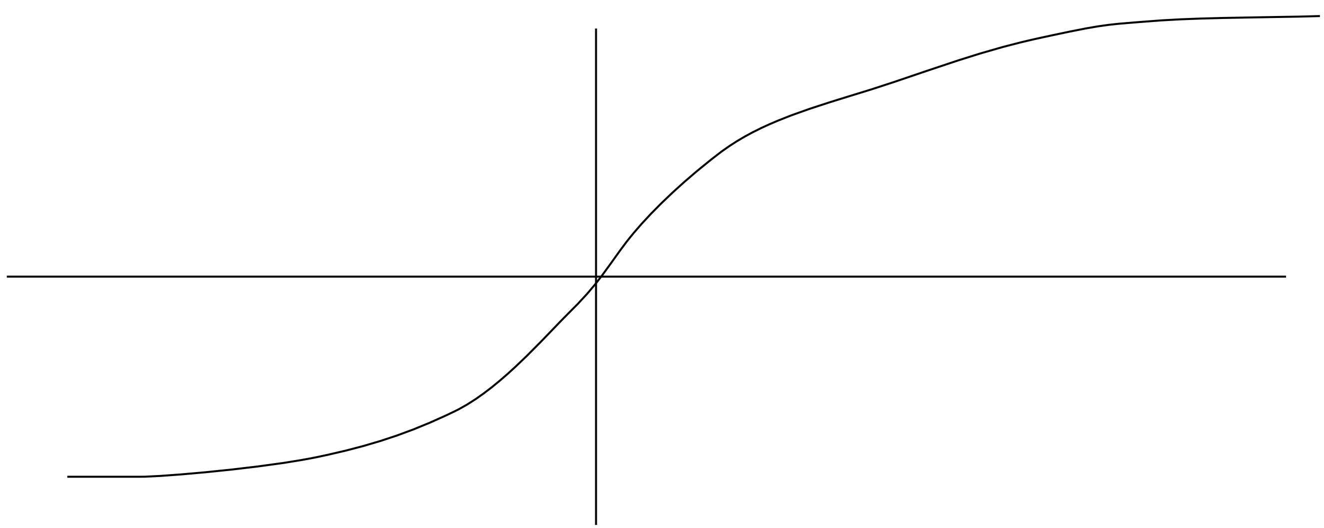
(e). This means that x_1 will always be larger than $c^{1/m}$ and that thereafter x_n will decrease monotonely towards $c^{1/m}$.

Exercise 2 (Divergence of Newton's method)
(Adapted from (Lange, 2010, exercise 5.12))
Assume we want to find the roots for the two functions

$$f(x) = \begin{cases} \sqrt{x} & x \geq 0 \\ -\sqrt{-x} & x < 0 \end{cases}$$

$$g(x) = x^{1/3}$$

What happens when you apply Newton's method in these cases?



(a). Consider first $f(x)$. Then

$$f'(x) = \begin{cases} \frac{1}{2\sqrt{x}} & x \geq 0 \\ \frac{1}{2\sqrt{-x}} & x < 0 \end{cases}$$

If $x_{n-1} > 0$ we get

$$x_n = x_{n-1} - \frac{\sqrt{x_{n-1}}}{\frac{1}{2\sqrt{x_{n-1}}}} = x_{n-1} - 2x_{n-1} = -x_{n-1}$$

while if $x_{n-1} > 0$ we get

$$x_n = x_{n-1} - \frac{\sqrt{-x_{n-1}}}{\frac{1}{2\sqrt{-x_{n-1}}}} = x_{n-1} - 2x_{n-1} = -x_{n-1}$$

so the algorithm will alternate between x_0 and $-x_0$.

(b). Consider now $g(x)$. Then

$$\begin{aligned} g'(x) &= \tfrac{1}{3}x^{-2/3} \\ x_n &= x_{n-1} - \frac{x_{n-1}^{1/3}}{\frac{1}{3}x_{n-1}^{-2/3}} \\ &= x_{n-1} - 3x_{n-1} = -2x_{n-1} \end{aligned}$$

so in this case the algorithm will diverge!

Ex3 f