UNIVERSITY OF CAMBRIDGE

MPhil in Population Health Sciences

Self-Screening Questions for Health Data Science

Version 0.9

A note from the admissions committee

This worksheet is for those considering an application to study the Health Data Science stream of the MPhil in Population Health Sciences at the University of Cambridge. Mathematical skills typical of an excellent A-level student are a minimum requirement for applicants to this stream. The questions are designed to allow the potential applicant to assess for her/himself whether s/he is likely to be able to meet the mathematical standard when s/he begins the course, if s/he is offered a place.

As you work through the questions, it is important that you are not put off if you find some of them challenging. We appreciate that many applicants, particularly those with a medical or biological background, may not have studied mathematics for some time. The aim is not to test whether you are ready to tackle the MPhil course today, but to give you the opportunity to assess whether you will be able to put yourself in the position to start the course with the level of mathematics required. If you find a question difficult, please look up the definitions of any terms of which you are unsure, study the resources suggested and return to the question. If you find you are unable to make progress after much further study you may wish to refer to the solutions, but we suggest that you consider doing so as a last resort. If you refer to the solutions, bear in mind that there may be other approaches to an answer that are correct. We hope you will view this as an opportunity to learn or review some mathematics, and, if you apply successfully, to ensure you are ready to take the course when you arrive in Cambridge.

If you have a strong interest in studying population health sciences but find many of these questions difficult (i.e. if you need to refer to many of the solutions even after further study) you may wish to consider applying to a stream with a less stringent mathematical requirement. For example, you could consider the Epidemiology stream, which is less mathematically demanding than the Health Data Science stream, but nevertheless provides a quantitative training and includes courses in biostatistics. A mathematical self-screening tool for the other streams is available from the the course website.

If you find an error please report to wja24@cam.ac.uk.

Question. Graphs

Each equation corresponds to a graph. Identify each correspondence and give your reasoning.

(a) $y = x + 1$

Solution: F

This is the equation of a line with x-axis intercept equal to -1 and y-axis intercept equal to 1 (i.e. the line passes through the points $(-1, 0)$, $(0, 1)$).

(b) $y = -x + 1$

Solution: B

This is the equation of a line with x-axis intercept equal to 1 and y-axis intercept equal to 1 (i.e. the line passes through the points $(1, 0)$, $(0, 1)$).

(c) $y = \frac{2}{3}x^2 - \frac{8}{3}x + 2$

Solution: E

$$\frac{2}{3} \cdot 1^2 - \frac{8}{3} \cdot 1 + \frac{6}{3} = 0,$$

so the quadratic equation has a root at 1 and factorises as:

$$y = \frac{2}{3}(x - 1)(x - 3).$$

This is the equation of a parabola passing through $(1, 0)$ and $(0, 2)$. The vertex is $(1, -\frac{2}{3})$.

(d) $y = -\frac{1}{30}(x - 1)(x + 61)$

Solution: G

This is the equation of a parabola with a root at 1 and y-axis intercept at $\frac{61}{30} \approx 2$. The standard form of the equation is:

$$y = -\frac{1}{30}x^2 - 2x + \frac{61}{30},$$

When x is small, this may be approximated as:

$$\begin{aligned} y &= -2x + \frac{61}{30} \\ &= -2x + 2. \end{aligned}$$

(e) $y = \frac{2^{16}}{3^6} x^6 (1-x)^2$

Solution: H

The equation has roots at 0 and 1, so its graph passes through the points $(0,0)$ and $(1,0)$. Because log is a strictly monotonic function, any stationary points of the function

$$f(x) = \frac{2^{16}}{3^6} x^6 (1-x)^2$$

are also stationary points of $\log f$. We can identify any such stationary points by differentiation:

$$\begin{aligned} 0 &= \frac{\mathrm{d} \log f}{\mathrm{d}x} \\ &= \frac{\mathrm{d}}{\mathrm{d}x} \left[6 \log x + 2 \log(1-x) \right] \\ &= \frac{6}{x} - \frac{2}{1-x}, \end{aligned}$$

which implies $x = \frac{3}{4}$. $f(3/4) = \frac{2^{16}}{3^6} (\frac{3}{4})^6 (\frac{1}{4})^2 = 1$ so the unique stationary point is at $(\frac{3}{4}, 1)$.

(f) $y = \frac{2^{16}}{3^6} x^2 (1-x)^6$

Solution: A

The solution is analogous to that of part (e).

(g) $y = \frac{x^2}{\log(e^{-x})} - 1$

Solution: I

$$\begin{aligned} y &= \frac{x^2}{\log(e^{-x})} - 1 \\ &= \frac{x^2}{-x} - 1 \\ &= -x - 1. \end{aligned}$$

This is the equation of a line with x-axis intercept equal to -1 and y-axis intercept equal to -1 (i.e. the line passes through the points $(-1, 0)$, $(0, -1)$).

(h) $y = \sin(4x^2 + \pi/2)$

Solution: D

$$\begin{aligned} y &= \sin(4x^2 + \pi/2) \\ &= \cos(4x^2). \end{aligned}$$

Consequently, the graph corresponds to an even function of x, with y-axis intercept 1. The equation has roots satisfying:

$$4x^2 = \frac{\pi}{2} + n\pi$$

that is,

$$x = \pm\frac{1}{2}\sqrt{\pi(\frac{1}{2} + n)},$$

for integer $n \geq 0$. There must be at least 8 roots in the interval $-2 < x < 2$, corresponding to the two choices of sign and $n = 0, 1, 2, 3$ because,

$$\begin{aligned} n < 3 &\implies n < 3 + \frac{1}{2} \\ &\implies \frac{1}{2} + n < 4 \\ &\implies 4\left(\frac{1}{2} + n\right) < 16 \\ &\implies \pi\left(\frac{1}{2} + n\right) < 16 \\ &\implies \sqrt{\pi\left(\frac{1}{2} + n\right)} < 4 \\ &\implies \frac{1}{2}\sqrt{\pi\left(\frac{1}{2} + n\right)} < 2. \end{aligned}$$

(In fact there are 10 roots in the interval, but this is enough to identify the correct graph.)

(i) $y = \cos(2\sqrt{|x|})$

Solution: C

The graph is of an even function with y-axis intercept equal to 1. The equation has roots satisfying

$$x = \pm\frac{1}{4}\left(\frac{\pi}{2} + n\pi\right)^2$$

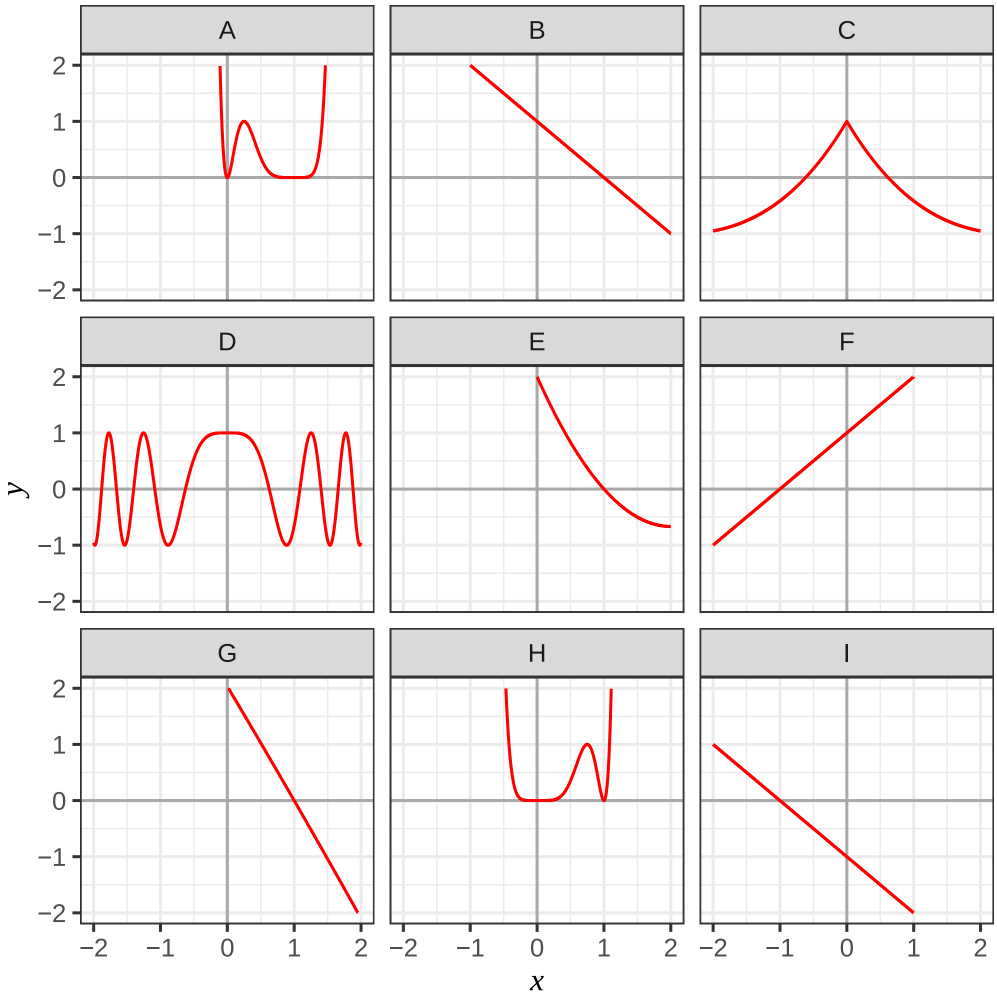
for integer $n \geq 0$. Because,

$$\frac{1}{4}\left(\frac{\pi}{2}\right)^2 < \frac{1}{4}\left(\frac{4}{2}\right)^2 = 1 < 2$$

and

$$2 < 5 < \frac{80}{16} < \frac{1}{4}\left(\frac{9}{2}\right)^2 < \frac{1}{4}\left(\frac{3}{2} + 3\right)^2 < \frac{1}{4}\left(\frac{\pi}{2} + \pi\right)^2$$

there are exactly two roots in the interval $-2 < x < 2$.



Question. Regions

Of the square in the (x, y) plane satisfying,

$$-5 < x < 5$$
$$-5 < y < 5,$$

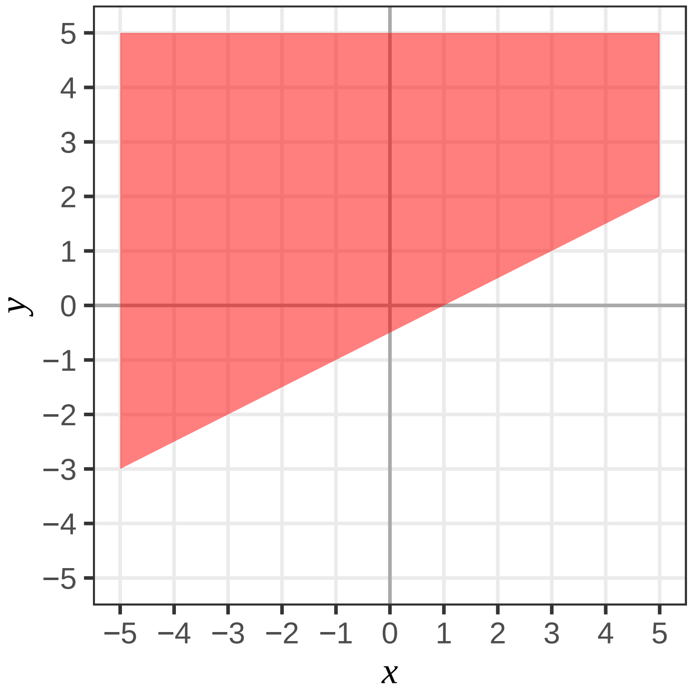
sketch the sub-region determined by each of the inequalities.

(a) $x < 2y + 1$

Solution:

The boundary internal to the square is determined by the equation $y = \frac{1}{2}x - \frac{1}{2}$. This corresponds to a line of gradient $\frac{1}{2}$ with y-axis intercept $-\frac{1}{2}$.

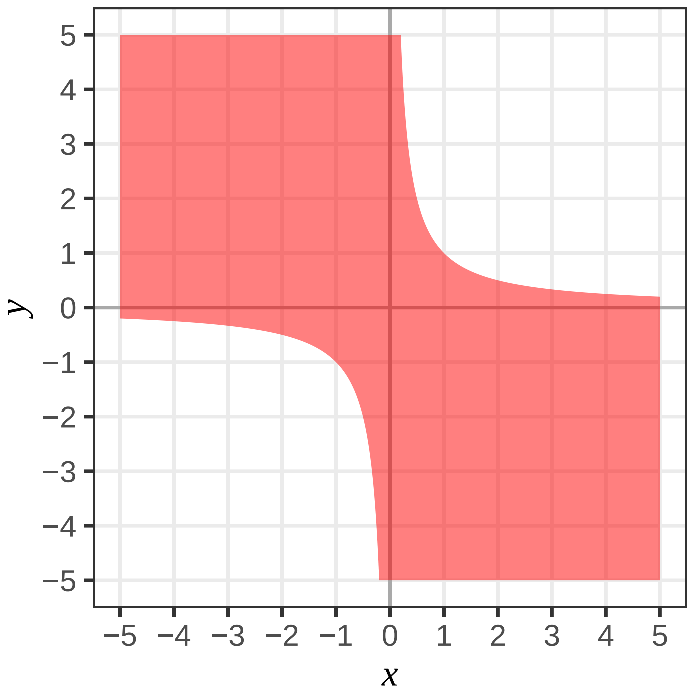
The region determined by the inequality lies above the line.



(b) $xy < 1$

Solution: There are two disconnected boundary components internal to the square, both determined by $y = \frac{1}{x}$. This is the equation of a hyperbola passing through $(1, 1)$ and $(-1, 1)$ with asymptotes $x = 0$ and $y = 0$.

The region determined by the inequality contains the origin $(0, 0)$.

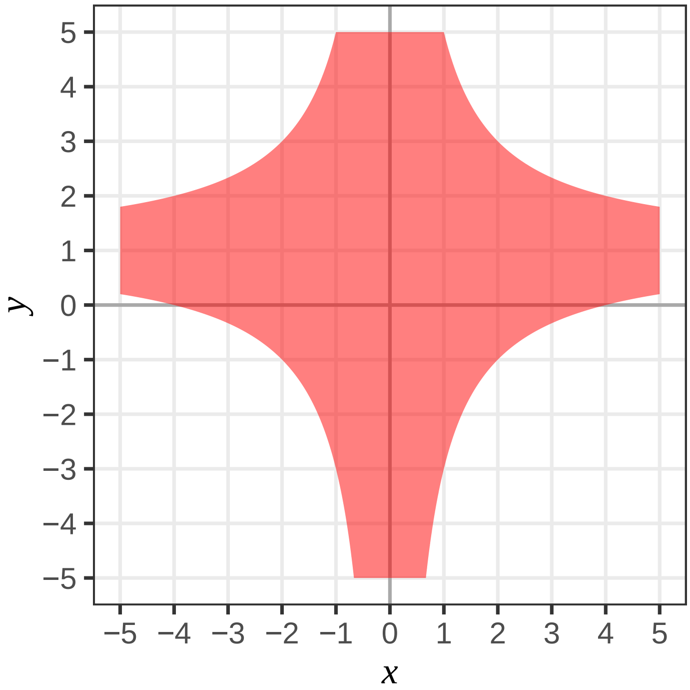


(c) $|x||y-1| < 4$

Solution: There are four disconnected boundary components internal to the square. If the sign of x is equal to the sign of $y-1$ then points on the internal boundary satisfy $y = 1 + \frac{4}{x}$. This is the equation of a hyperbola with asymptotes $x = 0$ and $y = 1$, one branch of which passes through $(1, 5)$, $(2, 3)$ and $(4, 2)$, the other branch of which passes through $(-1, -3)$, $(-2, -1)$ and $(-4, 0)$.

If the sign of x is different to the sign of $y-1$ then points on the internal boundary satisfy $y = 1 - \frac{4}{x}$. This is the equation of a hyperbola with asymptotes $x = 0$ and $y = 1$, one branch of which passes through $(1, -3)$, $(2, -1)$ and $(4, 0)$, the other branch of which passes through $(-1, 5)$, $(-2, 3)$ and $(-4, 2)$.

The region determined by the inequality contains the origin $(0, 0)$.

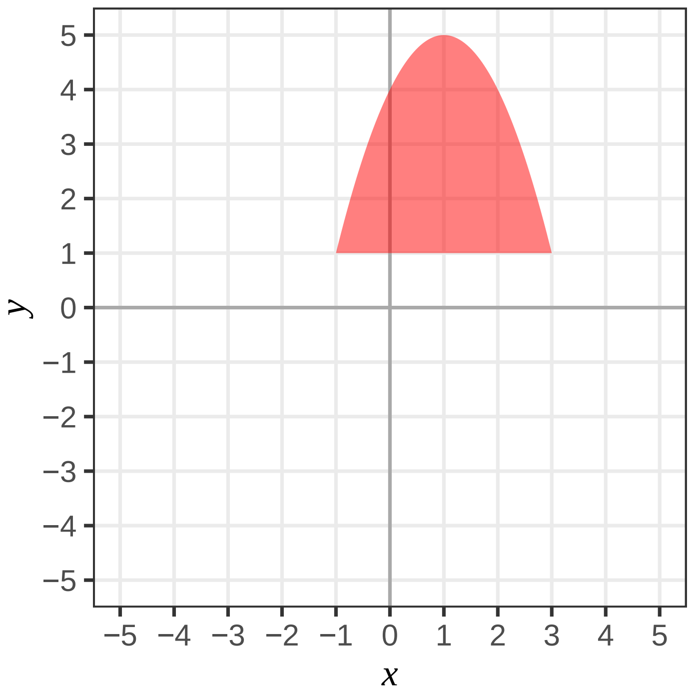


(d) $1 < y$ and $(x-2)x + y < 4$

Solution: Note that the quadratic inequality may be rewritten:

$$y < -(x+1)(x-3) + 1.$$

The vertex of the parabola corresponding to $y = -(x+1)(x-3)+1$ is at $(1, 5)$. Consequently, the boundary of region determined by the two inequalities is strictly inside the square except at the vertex point and consists of the segment of the parabola between $(-1, 1)$ and $(3, 1)$ and the segment of the line $y = 0$ between the same points.



(e) $(x+1)^2 + (y-2)^2 < 2|x+y-1| + 2$

Solution:

This is easier to tackle after the change of coordinates:

$$x' = x + 1$$
$$y' = y - 2.$$

The inequality in this system is

$$x'^2 + y'^2 < 2|x' + y'| + 2.$$

Now consider the cases $x' + y' \leq 0$ and $x' + y' > 0$ separately. In the first case, the inequality can be written

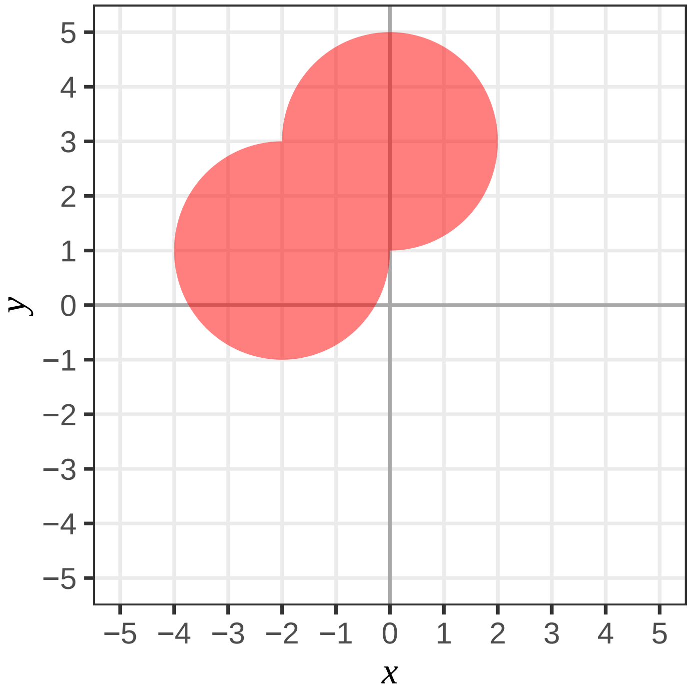
$$(x' + 1)^2 + (y' + 1)^2 < 2^2.$$

In the second case, the inequality can be written

$$(x' - 1)^2 + (y' - 1)^2 < 2^2.$$

Consequently, the region can be thought of as the disjoint union of a part of each of two discs: the part of the disc of radius 2 with centre $(x, y) = (-2, 1)$ that is below or on the line $y = -x + 1$ and the part of the disc of radius 2 with centre $(x, y) = (0, 3)$ that is above the line $y = -x + 1$.

Solution:



Question. Exponentials and Logarithms

The exponential function, $x \mapsto e^x$, is monotonic and strictly increasing. For every pair of numbers (a,b), it satisfies the identities:

$$e^a e^b = e^{a+b}$$
$$(e^a)^b = e^{ab}.$$

The natural logarithm function $x \mapsto \log x$ is the inverse of the exponential function, i.e. $\log x$ is such that $e^{\log x} = x$, for every $x > 0$.

Of the following statements, determine which are true and which are false for all $x > 0$ and all $y > 0$.

(a) $\dfrac{1}{(e^x)^y} = e^{-xy}$

Solution: True.

Note that:

$$\begin{aligned} 1 &= e^0 \\ &= e^{-xy+xy} \\ &= e^{-xy}e^{xy} \\ &= e^{-xy}(e^x)^y. \end{aligned}$$

Next, divide both sides by $(e^x)^y$ to obtain the result.

(b) $2e^{x-y} - 2e^{y-x} = (e^x - e^{-x})(e^y + e^{-y}) - (e^x + e^{-x})(e^y - e^{-y})$

Solution: True

$$\begin{aligned} 2e^{x-y} - 2e^{y-x} &= 2e^x e^{-y} - 2e^{-x}e^y \\ &= \left[e^x e^{-y} - e^{-x}e^y\right] - \left[e^{-x}e^y - e^x e^{-y}\right] \\ &= \left[e^x(e^y + e^{-y}) - e^x(e^y - e^{-y})\right] \dots \\ \dots &- \left[e^{-x}(e^y + e^{-y}) + e^{-x}(e^y - e^{-y})\right] \\ &= (e^x - e^{-x})(e^y + e^{-y}) - (e^x + e^{-x})(e^y - e^{-y}) \end{aligned}$$

(c) $\dfrac{e^{y+x-2}}{e^{x^2+x-6}} = e^{y-x+3}$

Solution: False.

For example, put $x = 2, y = 0$, then the left-hand side resolves to 1 and the right-hand side resolves to e^{-1}.

(d) $y\dfrac{e^x}{\sqrt{x}} = e^{\log y + x - \frac{1}{2}}$

Solution: False.

Put $x = 1, y = 1$, then the left-hand side resolves to e and the right-hand side resolves to $\sqrt{e}$.

(e) $y^{x+\frac{1}{8}} = e^{x \log y}\sqrt{\sqrt{\sqrt{y}}}$

Solution: True.

$$\begin{aligned} y^{x+\frac{1}{8}} &= \left(e^{\log y}\right)^{x+\frac{1}{8}} \\ &= e^{x \log y + \frac{\log y}{8}} \\ &= e^{x \log y} e^{\frac{\log y}{8}} \\ &= e^{x \log y} \left(e^{\log y}\right)^{\frac{1}{8}} \\ &= e^{x \log y} \left(\left(\left(e^{\log y}\right)^{\frac{1}{2}}\right)^{\frac{1}{2}}\right)^{\frac{1}{2}} \\ &= e^{x \log y}\sqrt{\sqrt{\sqrt{y}}} \end{aligned}$$

(f) $\frac{x}{y} = e^{\log_y x}$

Solution: False.

Put $x = 1, y = e$, then the left-hand side resolves to e^{-1} and the right-hand side resolves to 1.

(g) $\log(x + y) = \log x \log y$

Solution: False.

Put $x = 1, y = 1$. With this choice, the left-hand side of the equation is equal to $\log 2$. $\log 2 > \log 1 = 0$, because log is a strictly increasing monotonic function. On the other hand, the right-hand side of the equation is equal to $(\log 1)^2 = 0$. This is a contradiction.

(h) $\log x + \log y = \log(xy)$

Solution: True.

This is a standard logarithmic identity, but it can be derived from the exponential identity $e^a e^b = e^{a+b}$. log is the inverse of exp, consequently,

$xy = e^{\log(xy)}$ and therefore,

$$\begin{aligned} e^{\log(xy)} &= xy \\ &= e^{\log x} e^{\log y} \\ &= e^{\log x + \log y}, \end{aligned}$$

from which the result follows, by taking logs on both sides.

(i) $x^{\log y} = y^{\log x}$

Solution: True.

$$\begin{aligned} x^{\log y} &= (e^{\log x})^{\log y} \\ &= e^{\log x \log y}, \end{aligned}$$

from which the result follows, because of the symmetry between x and y on the right-hand side.

(j) $\log_x(y) = \dfrac{\log x}{\log y}$

Solution: False.

Put $x = e$, in which case the assertion is equivalent to:

$$\begin{aligned} (\log y)^2 &= \log e \\ &= 1. \end{aligned}$$

Put $y = \frac{1+e}{2}$, then, because log is monotonic increasing,

$$1 < y < e \implies 0 < \log y < 1 \implies (\log(y))^2 < \log(y) < 1,$$

a contradiction.

(k) $\log_{1/x}(y) = -\dfrac{\log y}{\log x}$

Solution: True

$$\begin{aligned} y &= e^{\log y} \\ &= (e^{\log x})^{\frac{\log y}{\log x}} \\ &= x^{\frac{\log y}{\log x}} \\ &= \left(\frac{1}{x}\right)^{-\frac{\log y}{\log x}}. \end{aligned}$$

The result follows because $\log_{1/x}$ is defined as the inverse of $z \mapsto (\frac{1}{x})^z$.