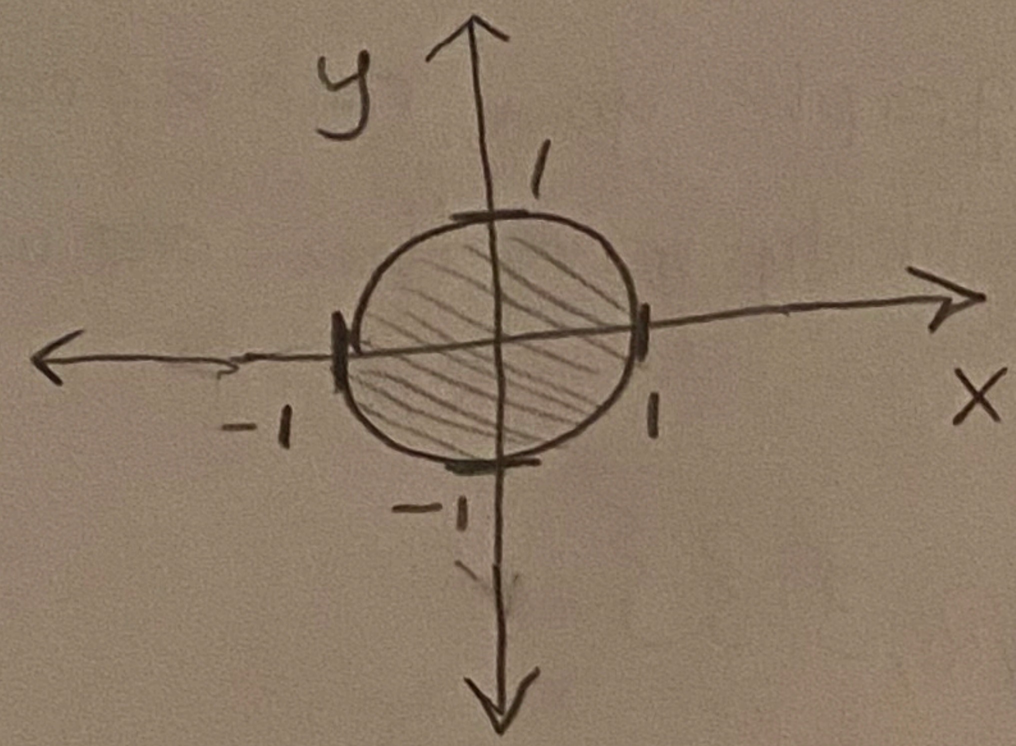
Supplementary Review Problems and Comments

1. Math symbols & Nomenclature

Draw $\{(x,y) \in \mathbb{R}^2 \mid x^2+y^2 \leq 1\}$ note $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ means 2-d
points/vectors $\rightarrow \mathbb{R}^3$ 3-d
$\mathbb{R}^n$ n-dimensional

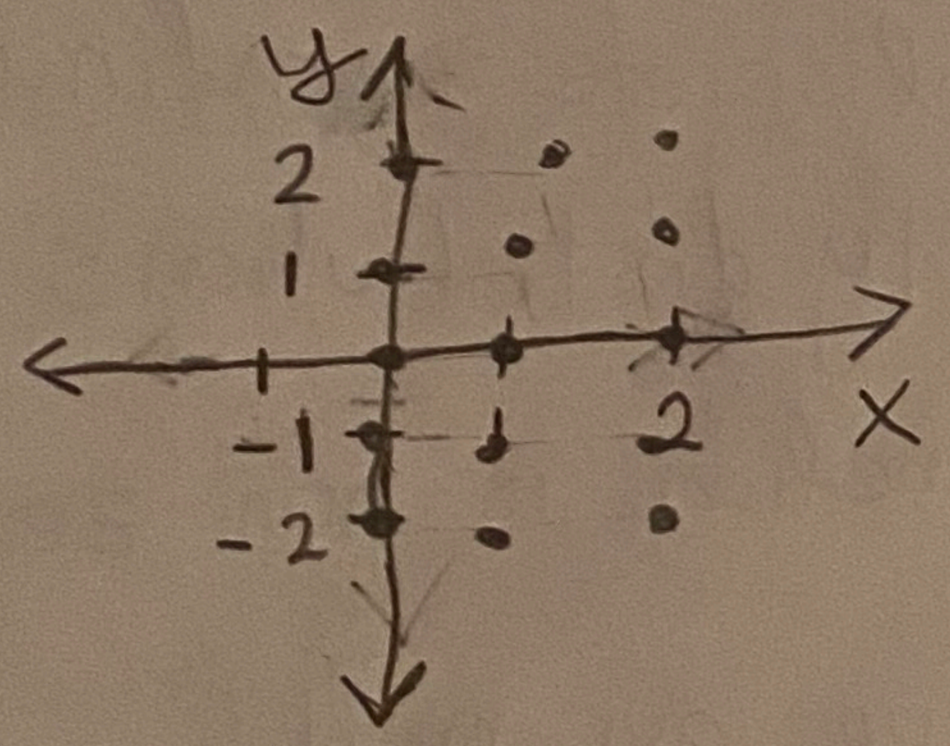
Answer:



region within unit circle

Draw $\{(x,y) \in \mathbb{Z}^2 \mid 0 \leq x \leq 2, -3 < y < 3\}$

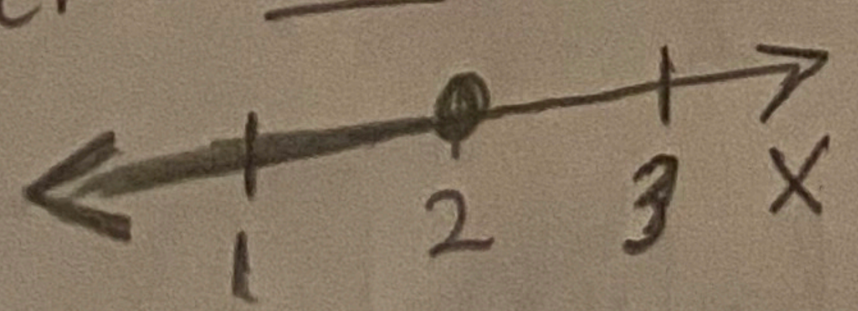
Answer:

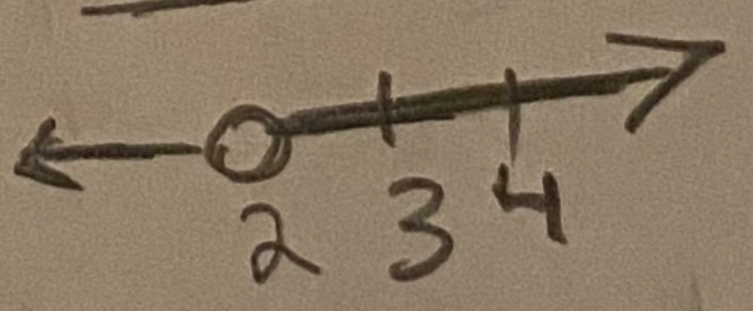


only integer pairs are valid
$-3 < y < 3$ for discrete integers is equivalent to
$-2 \leq y \leq 2$

Draw A^c where "c" stands for complement and
$A = \{x \in \mathbb{R} \mid x \leq 2\}$
Bonus: write the set builder notation for A^c

Answer: A ray to the left $\quad A^c$ ray to the right $\{x \in \mathbb{R} \mid x > 2\}$

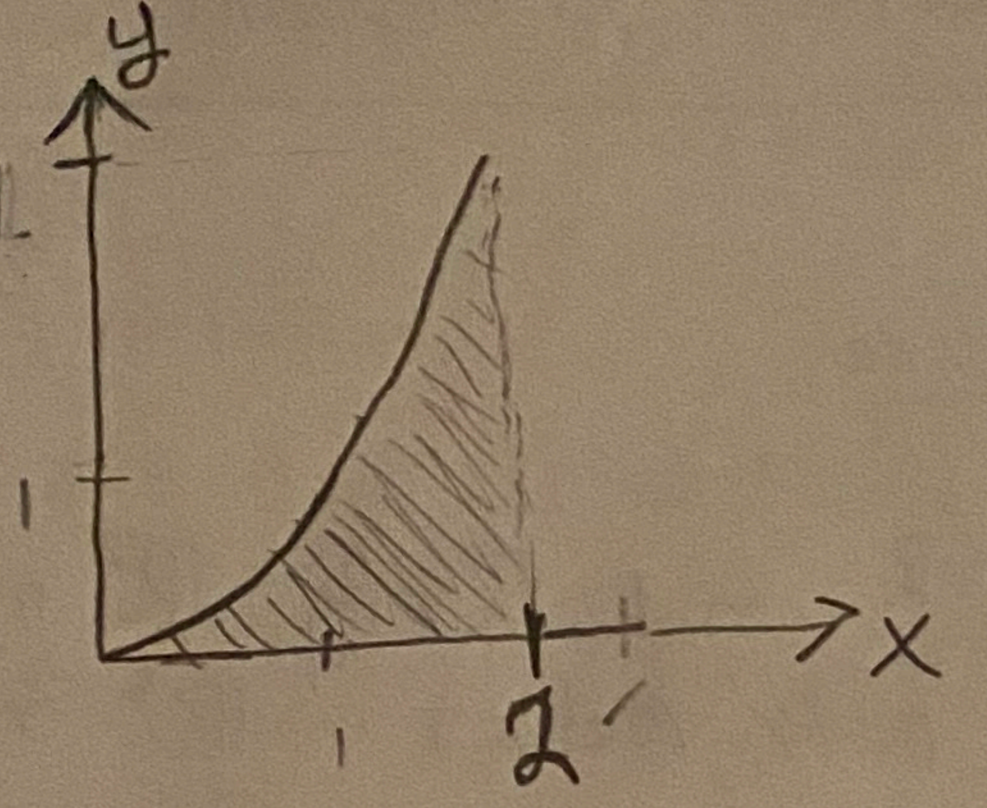




What about $A \cup A^c$?
$A \cap A^c$?

Draw: $\{(x,y) \in \mathbb{R}^2 \mid 0 \le x^2 \le y \le 4\}$

Answer:



Try: see if you can translate graphs you run across into sets or events. There is likely more than one way to describe them.

Eg. set of odd rolls on a die

$\{x \mid x = 2n-1 \text{ where } n \in \{1,2,3\}\}$

$\{1,3,5\}$

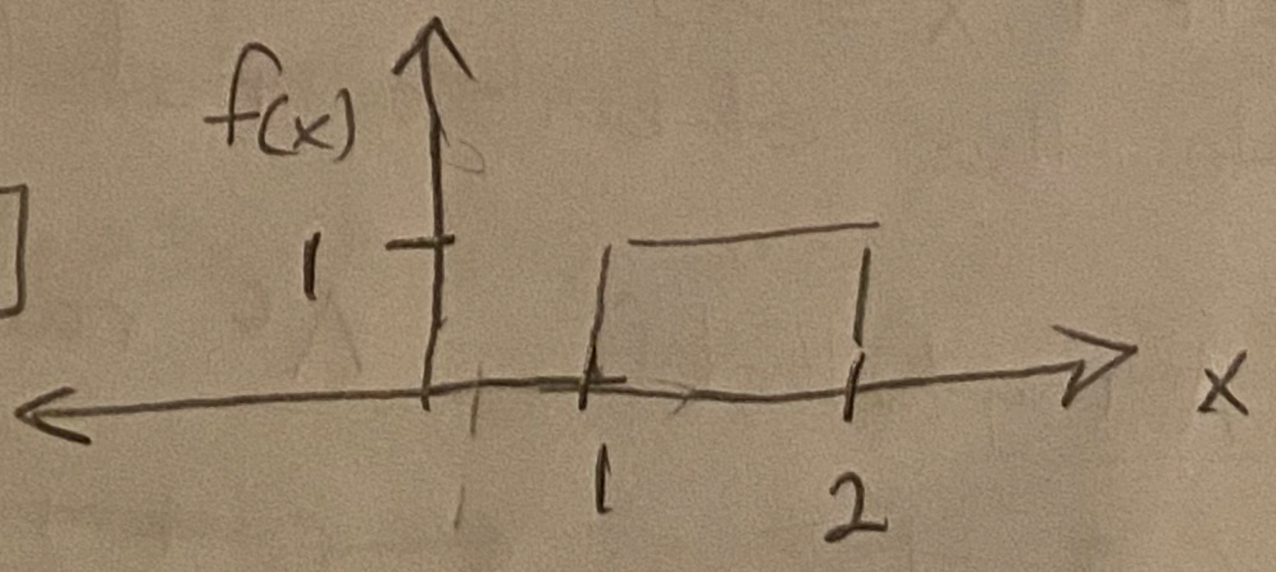
Sets show up all across probability in discrete & continuous cases. An important set for probability distributions the support, the region where our function is non-zero

This is helpful for simplifying & setting up our math, especially when integrating.

Consider a uniform distribution over the interval $[1,2]$

ie: $X \sim U(1,2)$ such that

$f(x) = 1$ for $x \in [1,2]$
$\quad\;\; 0$ else



random variable X distributed as uniform over $[1,2]$

If we ask to compute the average/expectation $E[X]$ formally, we have $E[X] = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$

But really $E[X] = \int_{\text{support}(X)} x \cdot f(x)\,dx = \int_1^2 x \cdot 1\,dx$ which is tractable

Set up $\int_{\text{Support}(f(x))} f(x)\,dx$ where $f(x) = \begin{cases} 2x & -5 < x < -2 \\ -2x+2 & -2 \le x < 2 \\ 0 & \text{else} \end{cases}$

Answer: $\int_{-5}^{-2} 2x\,dx + \int_{-2}^{2} -2x+2\,dx$

2 Logarithms & Exponents

For this class, we mainly deal with natural log (base $e \approx 2.7$) ln

A few key things beyond the rules mentioned on the review sheet

- log is a monotonic transformation (i.e if $x<y$, $f(x)<f(y)$) that preserves location of extrema
  (ie if 5 is where $f(x)$ is max, $\ln(f(x))$ also maximizes at $x=5$)
- simplifies sums to probability
  $\ln(f(x) \cdot g(x) \cdot h(x)) = \ln(f) + \ln(g) + \ln(h)$
  easier to take derivative of, also numerically stable (multiplying many small numbers quickly hit below computer threshold & effictively equal 0 vs adding small numbers to preserve magnitude)

Important for maximum likelihood estimation when we get there, regression, machine learning algorithms like Naive Bayes

- ln & e are inverses → $e^{\ln(x)} = x$

Simplify

- $e^{\ln(x^2)}$
- $e^{2x} e^{5x+3\ln(x)}$
- $\ln(2x) + 3\ln(x) + \ln(x)$

Answers:

- x^2
- $x^3 e^{7x}$
- $\ln(2x \cdot x^3 \cdot x) = \ln(2x^5)$

3 Discrete Data & Summation Notation

e.g. $1+2+3+4+\dots 10 = \sum_{i=1}^{10} i$

Product notation:

e.g. $1\cdot 2\cdot 3\cdot 4\cdot \dots \cdot 10 = \prod_{i=1}^{10} i$

note: $\ln(\prod_i) = \sum \ln(i)$

We also have infinite series

e.g. $\sum_{n=1}^{\infty} 1/n^2$ or $\sum_{k=1}^{\infty} 3^k$ or $\sum (-1)^n \frac{2n+1}{3n+2}$

have notions of convergence (sums to a finite number)
and divergence (does not sum/settle to a number)

Some common ones with names are power series, geometric, Taylor, etc
we don't touch these much in this class but good to be aware of
—(at least not explicitly...)

" Σ " will be our discrete analogy to " $\int$ "
summations integrals

4 Functions & Functional Notation

We can add, subtract, compose functions, output is still a function

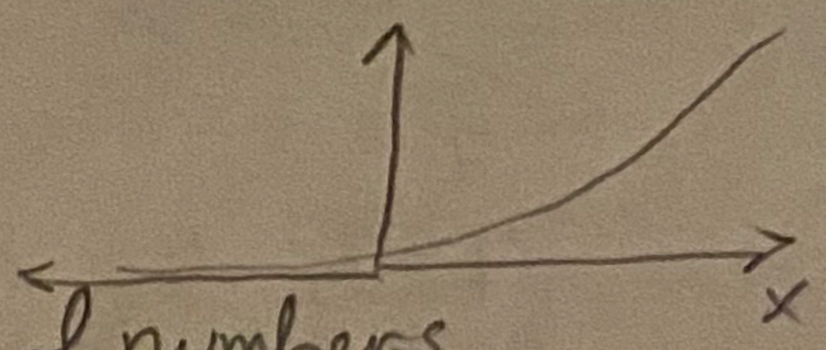
- $h = f+g$ e.g. $f = 2x$ $g = \cos 2x + 3x + 1$ $h(x) = \cos 2x\ 5x+1$
- $h = f-g$ $f = 5x^2$ $g = 4x^2 + e^x$ $h(x) = x^2 - e^x$
- $h = f\circ g =$ $f = e^{2x}$ $g = x^2+1$ $h(x)\ e^{2\cdot(x^2+1)}$

$\overset{\text{ll}}{f(g(x))}$
input to g
input to f

4 cont. Functions

- note: to properly define function, need to specify for all values of x

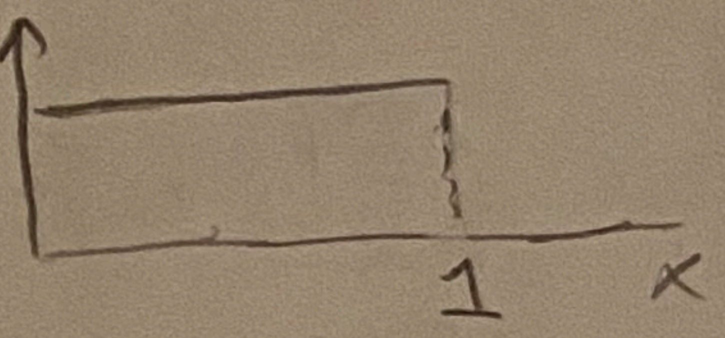
e.g $f(x) = e^x, \forall x \in \mathbb{R}$



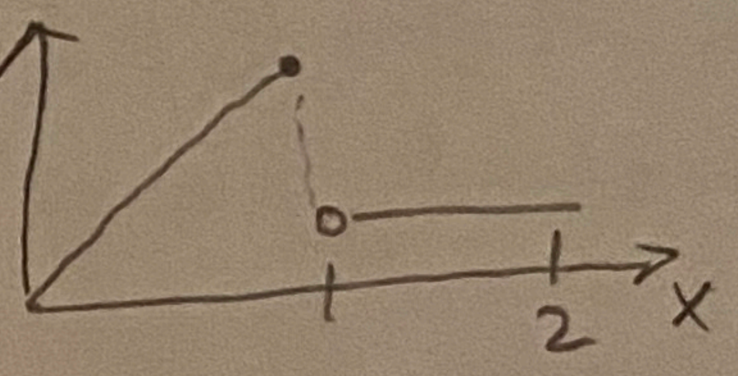
for all real numbers

commonly for pdf's/pmfs we will have piecewise functions:

e.g. $f(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$



e.g. $f(x) = \begin{cases} x & 0 \le x \le 1 \\ \frac{1}{2} & 1 < x < 2 \\ 0 & \text{otherwise} \end{cases}$



4 cont.

5/6 derivatives & integrals

common functions: e^x, $\ln(x)$, $ax^n + bx^{n-1} + \dots$ — polynomials

compositions, addition, and subtraction of these is also fair game

e.g. for a bell curve / Guassian, function is roughly e^{-x^2}

we don't really touch trignometric fxns in this class

expected rules to be familiar with (part 2 of review sheet)

$P(X=x)$ pmf — "Σ" — cmf $P(X \le x)$

$f(x)$ pdf — "$\int$" — cdf — $\frac{d}{dx}$

look at the "jumps" in probability

derivatives
- product rule
- chain rule
- power rule (polynomials)
- derivatives of sums / diffs

integrals
- u-sub
- integration by parts
- "reverse" power rule (polynomials)

Fundamental Thm

- If $F(x) = \int_a^x f(t)\,dt$
  $F'(x) = \frac{d}{dx} F(x) = f(x)$ → derivative & integral "undo" one another
- $\int_a^b f(t)\,dt = F(b) - F(a)$ → if the indefinite integral $\int f$ is known, the definite only cares about endpoints

there are assumptions on continuity & differentiability, for this class, our functions are "nice enough" to obey the rules

Note: Do not apply these rules to discrete functions. May seem obvious but can be forgotten in the moment. Remember "$\int$" & "Σ" are analogous but NOT identical.

Linearity of derivatives

$\frac{d}{dx}(a \cdot f(x) + g(x))$, $a \in \mathbb{R}$

$= a\frac{df}{dx} + \frac{dg}{dx}$

Power Rule

$\frac{d}{dx} x^n = nx^{n-1}$

Eg. $\frac{d}{dx}(-3x^2+5x) = -3\frac{d}{dx}(x^2) + 5\frac{d}{dx}(x)$

$= -3 \cdot (2)x^{2-1} + 5 \cdot x^{1-1}$

$= -6x + 5$

$\frac{d}{dx} e^x = e^x$ $\quad \frac{d}{dx} \ln(x) = \frac{1}{x}$

Chain Rule

$\frac{d}{dx} h(x) = \frac{d}{dx}(f(g(x))) = \frac{df}{dx}(g) \cdot \frac{dg}{dx}$

Eg $\underbrace{e^{x^2+2}}_{h}$ $\quad f = e^x$ $\quad \rightarrow \frac{d}{dx} e^{x^2+2} = e^{x^2+2} \cdot 2x$

$\underbrace{g = x^2+2}_{\text{power rule}}$

Eg $\ln(2x)$ $\quad \rightarrow \frac{d}{dx}\ln(2x) = \frac{1}{2x} \cdot 2 = \frac{1}{x}$

$f = \ln(x)$

$g = 2x$

Product Rule

$\frac{d}{dx}(f(x) \cdot g(x)) = \frac{df}{dx} \cdot g + \frac{dg}{dx} \cdot f$

Eg $x^2 e^{3x} \xrightarrow{d/dx} 2xe^x + 3x^2 e^{3x}$ — Chain rule of e^{3x} as well

Finding Extrema (Max/Min)

For $f(x)$, to find max/min

set $\frac{df}{dx} = 0$ and for value(s)

check sign of $\underbrace{\frac{d^2 f}{dx}}_{\text{Curvature}}$ — second derivative / deriv. of deriv.

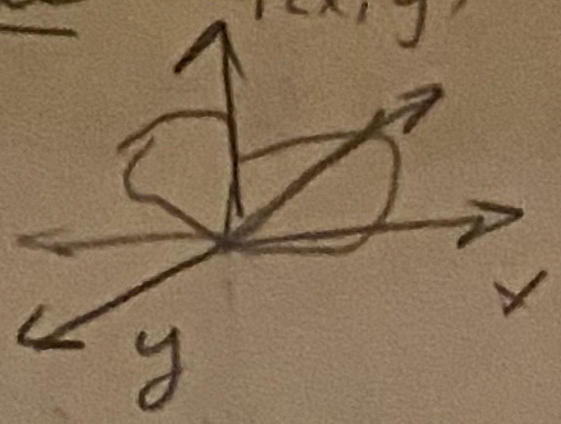
>0 concave up $\cup$ min

<0 concave down $\cap$ max

Multivariable

$f(x,y)$

$f(x,y)$



can now take the derivative with respect to each variable

→ partial derivatives, mechanically same rules apply

$$\text{e.g. } f(x,y) = x^2 + e^{2x}y^2 + 3yx^3$$

treat y as constant → $\dfrac{\partial f}{\partial x} = 2x + 2e^{2x}y^2 + 9yx^2$

x as constant → $\dfrac{\partial f}{\partial y} = 0 + 2e^{2x}y + 3x^3 = 2e^{2x}y + 3x^3$

Integrals

Reverse Pwr Rule

$$\int x^n\,dx = \frac{x^{n+1}}{n+1} + c$$

eg $x^3 + 2x^2 + 4 \xrightarrow{\int} \dfrac{x^4}{4} + \dfrac{2}{3}x^2 + 4x + c$

U-sub "reverse chain rule"

$$\int f(g(x))g'(x)dx = \int f(u)du = F(u) + c$$

where $u = g(x)$ $du = g'(x)$

eg. $\displaystyle\int \frac{x}{x^2+1}\,dx$ $\quad u = x^2+1$

$du = 2x\,dx \rightarrow \frac{1}{2}du = x\,dx$

$$\int \frac{1}{u}\cdot\frac{1}{2}\cdot du = \frac{1}{2}\ln(u) + c$$

$$= \frac{1}{2}\ln(x^2+1) + c$$

Integration By Parts "Reverse Product Rule"

$$\int v\frac{du}{dx}dx = uv - \int u\frac{dv}{dx}dx$$

eg $\int xe^x dx$ $\quad u = x \quad v' = e^x$, $u' = 1 \quad v = e^x$ $\rightarrow xe^x - \int e^x dx = xe^x - e^x + C$

Can apply these multiple times, in combination... Same for derivatives

General Notes on functions, recognizing output as fxn vs value

Important to recognize when answer/output should be a function or a value

eg. indefinite integral vs definite

$\underbrace{\int f(x)\,dx}_{\text{fxn}}$ vs $\underbrace{\int_0^1 f(x)\,dx}_{\text{value}}$

derivative

$\underbrace{\frac{d}{dx} f(x)}_{\text{fxn}}$

derivative at value

$\left.\frac{d}{dx} f(x)\right|_{x=3}$

value

finding extrema

$\frac{d}{dx} f(x) = 0$

could be either

generally a value

expected value

$E(X)$

value

vs

conditional expectation

$E(X|Y)$

function