MATH7501: Mathematics for Data Science I

Unit 9: Integration

Motivation: The need to integrate

The integral is, in some sense, the opposite of a derivative. When you take derivatives, you seek to study a function by looking at its 'infinitesimal' changes, whereas when you integrate, you seek to combine all of these infinitesimal changes to recover information about the original function.

For example, if you were driving with no GPS and no road signs, you could 'integrate' the data given to you by your speedometer to work out how far you have travelled.

Integration is **essential** in probability theory.

9.1 Area under a curve

Consider a positive and continuous function $f : [a, b] \to \mathbb{R}^+$.

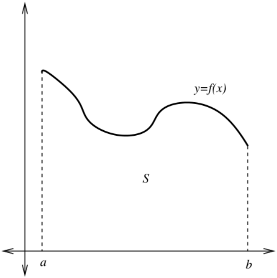


Figure 56: We need to define the area of the region $S = \{(x, y)|a \leq x \leq b, 0 \leq y \leq f(x)\}$.

How does one rigorously define the area underneath the curve $y = f(x)$?

9.1 Area under a curve

Divide the interval $[a, b]$ into subintervals $[x_0, x_1], [x_1, x_2], \cdots, [x_{n-1}, x_n]$, with $a = x_0$, $b = x_n$ and $x_0 < x_1 < \cdots < x_n$. For each i, choose $c_i \in [x_{i-1}, x_i]$. Then the area of S that lies between $[x_{i-1}, x_i]$ will be approximately $f(c_i) \times (x_i - x_{i-1})$.

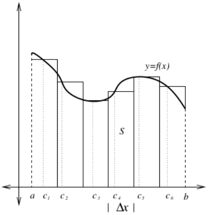


Figure 57: A method of approximating the area of the region S.

Then the total area will be approximately $\sum_{i=1}^{n} f(c_i)(x_i - x_{i-1})$. This approximation is called a *Riemann sum*.

9.1 Area under a curve

Definition (Area under a Curve)

The *area* under a positive and continuous function $f : [a, b] \to \mathbb{R}^+$ is defined to be

$$\lim_{n\to\infty} \sum_{i=1}^{n} f(x_i)(x_i - x_{i-1}),$$

as long as the size of the biggest sub-interval $[x_{i-1}, x_i]$ goes to 0 as n goes to ∞.

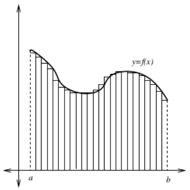


Figure 58: The smaller the width of the rectangle, the better the approximation, in general. Notice that in this diagram we chose $c_i = a + i\Delta x$.

9.1 Area under a curve

Definition (Riemann Integral)

For *any* continuous function $f : [a, b] \to \mathbb{R}$, the expression

$$\lim_{n\to\infty} \sum_{i=1}^{n} f(x_i)(x_i - x_{i-1})$$

from before is known as the *Riemann integral of f over* $[a, b]$, and is denoted $\int_a^b f(x)dx$
(although, the x in $f(x)dx$ can be replaced with any other 'dummy' variable).

If f is not positive on $[a, b]$, then the integral cannot be interpreted as the area under the curve. Rather, it gives a 'signed area', meaning

$$\int_a^b f(x)dx = \textit{area above x-axis, below graph} - \textit{area below x-axis, above graph}$$

9.1 Area under a curve

Definition (Anti-Derivative)

Let $f : [a, b] \to \mathbb{R}$ be a continuous function. An *antiderivative* for f is a function $F : [a, b] \to \mathbb{R}$ so that $F'(x) = f(x)$ for all $x \in (a, b)$.

Theorem (Fundamental Theorem of Calculus)

For a continuous function $f : [a, b] \to \mathbb{R}$, the function

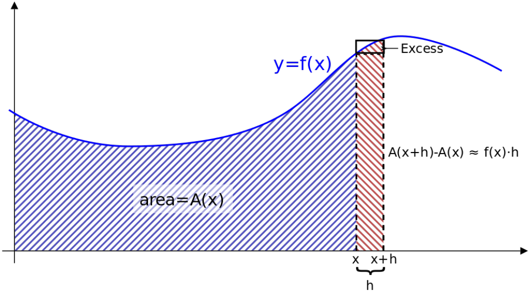
$$A(x) = \int_a^x f(t)dt$$

is an anti-derivative for f. If F is *any* anti-derivative for f, then

$$F(b) - F(a) = \int_a^b f(x)dx.$$

As notation, we often use $F(b) - F(a) = F(x)|_a^b = [F(x)]_a^b$.

9.1 Area under a curve



https://en.wikipedia.org/wiki/Fundamental_theorem_of_calculus

Theorem (Properties of Integrals)

Let c be a real number, and consider continuous functions $f, g : [a, b] \to \mathbb{R}$. Then

- $\int_b^a f(x)dx = -\int_a^b f(x)dx$ (this is more of a definition);
- $\int_a^b c dx = c(b - a)$;
- $\int_a^b (f(x) \pm g(x))dx = \int_a^b f(x)dx \pm \int_a^b g(x)dx$;
- $\int_a^b cf(x)dx = c\int_a^b f(x)dx$;
- $\int_a^c f(x)dx + \int_c^b f(x)dx = \int_a^b f(x)dx$;
- $\int_a^a f(x)dx = 0$ (again, sort of a definition).

9.1 Area under a curve

Question

Show that $\ln(x) = \int_1^x \frac{1}{t} dt$ for all $x > 0$.

Solution

Consider the function $g(x) = \ln(x) - \int_1^x \frac{1}{t} dt$. Using the properties of derivatives and the FTC, we find

$$g'(x) = \frac{1}{x} - \frac{1}{x} = 0.$$

The only functions with derivative 0 are constant (MVT). We therefore find that for all $x > 0$,

$$g(x) = g(1) = \ln(1) - \int_1^1 \frac{1}{t} dt = 0.$$

Question

Evaluate $\int_0^{\frac{\pi}{2}} \sin(x)dx$.

Solution

The function $-\cos(x)$ is an antiderivative for $\sin(x)$, so

$$\begin{aligned}\int_0^{\frac{\pi}{2}} \sin(x)dx &= [-\cos(x)]|_0^{\frac{\pi}{2}} \\ &= -\cos(\frac{\pi}{2}) - (-\cos(0)) \\ &= 1.\end{aligned}$$

9.1 Area under a curve

Question

Let $F(u) = \int_2^u \frac{dt}{\sqrt{1+t^2}}$. Find $F'(u)$.

Solution

By the FTC,

$$F'(u) = \frac{1}{\sqrt{1+u^2}}.$$

Question

Let $G(x) = \int_2^{\sin(x)} \frac{dt}{\sqrt{1+t^2}}$. Find $G'(x)$.

Solution

We can write $G(x) = F(\sin(x))$, where F is from the previous question. We already know F', so by the chain rule,

$$\begin{aligned} G'(x) &= F'(\sin(x))\cos(x) \\ &= \frac{\cos(x)}{\sqrt{1+\sin^2(x)}} \end{aligned}$$

9.1 Area under a curve

Question

Find the area between the x-axis and the curve $y = \sin(x)$ for $0 \leq x \leq \pi$.

Solution

The function $\sin(x)$ is non-negative on $[0, \pi]$, so the area is

$$\begin{aligned}\int_0^{\frac{\pi}{2}} \sin(x) dx &= [-\cos(x)]|_0^{\pi} \\ &= -\cos(\pi) - (-\cos(0)) \\ &= 2.\end{aligned}$$

9.1 Area under a curve

Question

Find the area bounded by the two curves $y = x^2$ and $y = x^3$ on the interval $[0, 1]$.

Solution

On the interval $[0, 1]$, we have $x^2 \geq x^3$, so the area between the two is the area below x^2, subtract the area below x^3. We compute

$$
\begin{aligned}
\text{area} &= \int_0^1 x^2 dx - \int_0^1 x^3 dx \\
&= \left[\frac{x^3}{3}\right] 0^1 - \left[\frac{x^4}{4}\right]_0^1 \\
&= \frac{1}{3} - \frac{1}{4} \\
&= \frac{1}{12}.
\end{aligned}
$$

9.1 Area under a curve

Corollary (of the FTC)

If f is non-negative and continuous on $[a, b]$, then $\int_a^b f(x)dx \geq 0$.

9.2 Approximate Integration

Let $f : [a, b] \to \mathbb{R}$ be a continuous function, and choose $a = x_0 < x_1 < \cdots < x_n = b$. One way to approximate $\int_a^b f(x)dx$ is with the left endpoint approximation:

$$\int_a^b f(x)dx \approx \sum_{i=1}^{n} f(x_{i-1})(x_i - x_{i-1}).$$

Another way is the right endpoint approximation:

$$\int_a^b f(x)dx \approx \sum_{i=1}^{n} f(x_i)(x_i - x_{i-1}).$$

Another way is the *trapezoidal rule*, which averages the previous two:

$$\int_a^b f(x)dx \approx \sum_{i=1}^{n} \frac{f(x_i) + f(x_{i-1})}{2}(x_i - x_{i-1}).$$

9.3 Indefinite Integrals

Definition (Indefinite Integrals)

Let f be some function. The *indefinite integral* of f, denoted $\int f(x)dx$, is the set of all anti-derivatives of f.

By the fundamental theorem of calculus, anti-derivatives only differ by a constant, so if F is *some* anti-derivative for f, it is common to write

$$\int f(x) = F(x) + c$$

for $c \in \mathbb{R}$.

9.3 Indefinite Integrals

Recall that if f is defined on an interval I, then an anti-derivative for f is a function F so that $F'(x) = f(x)$. Let c be some constant, f, g be some functions, and F, G some anti-derivatives for f, g. Some common anti-derivatives are given below:

Function	Antiderivative
$cf(x)$	$cF(x)$
$f(x)+g(x)$	$F(x)+G(x)$
$x^\alpha, (\alpha \neq -1)$	$\frac{x^{\alpha+1}}{\alpha+1}$
$\sin x$	$-\cos x$

Function	Antiderivative
$\cos x$	$\sin x$
$\sec^2 x$	$\tan x$
$\frac{1}{x}$	$\ln x$
e^x	e^x

But recall, the set of *all* anti-derivatives is found by adding constants to those in the table.

9.3 Indefinite Integrals

Question

Find $\int(x^2 + 3x)dx$, meaning, find all antiderivatives of the function $f(x) = x^2 + 3x$.

Solution

Using the table, we find that

$$\int x^2 + 3xdx = \frac{x^3}{3} + \frac{3x^2}{2} + C$$

for $C \in \mathbb{R}$.

Question

Suppose $f''(x) = x - \sqrt{x}$. Find $f(x)$.

Solution

Using the table, we find that $f'(x) = \frac{x^2}{2} - \frac{2x^{\frac{3}{2}}}{3} + C$ for some $C \in \mathbb{R}$.
Then using the table again,

$$f(x) = \frac{x^3}{6} - \frac{4x^{\frac{5}{2}}}{15} + Cx + D$$

for some $C, D \in \mathbb{R}$.

9.4 Explicit Integration Techniques

If possible, it is often handy to solve integrals explicitly. There are a number of standard tricks you should know. The first is the integral version of the chain rule:

Integration by Substitution

If $u = g(x)$ is a differentiable function with range in an interval I and f is continuous on I, then

$$\int f(g(x))g'(x)dx = \int f(u)\frac{du}{dt}dt = \int f(u).$$

If g' is continuous on $[a, b]$, then

$$\int_a^b f(g(x))g'(x)dx = \int_{g(a)}^{g(b)} f(u)du.$$

Question

Find $\int x^3 \cos(x^4 + 2) dx$.

Solution

Let $u = x^4 + 2$, so that $x^3 = \frac{1}{4}\frac{du}{dx}$, and

$$\begin{aligned} \int x^3 \cos(x^4 + 2) dx &= \int \frac{\cos(u)}{4} du \\ &= \frac{\sin(u)}{4} + C \\ &= \frac{\sin(x^4 + 2)}{4} + C. \end{aligned}$$

Question

Evaluate $\int_1^2 \frac{dx}{(5x-3)^2}$.

Solution

Substitute $u = g(x) = 5x - 3$, so that $\frac{du}{dx} = 5$, and

$$\begin{aligned}
\int_1^2 \frac{dx}{(5x-3)^2} &= \int_{g(1)}^{g(2)} \frac{du}{5u^2} \\
&= \frac{1}{5} \int_2^7 \frac{1}{u^2} du \\
&= \frac{1}{5} \Big[-u^{-1} \Big]_2^7 \\
&= \frac{1}{5} \left(-\frac{1}{7} + \frac{1}{2} \right) = \frac{1}{14}.
\end{aligned}$$

9.4 Explicit Integration Techniques

Question

Find $\int \frac{dx}{\sqrt{a^2-x^2}}$ for any $a > 0$.

Solution

Substitute $x = a\sin(\theta)$, so that $dx = a\cos(\theta)d\theta$. Then

$$\begin{aligned}
\int \frac{1}{\sqrt{a^2-x^2}}dx &= \int \frac{a\cos(\theta)d\theta}{\sqrt{a^2-a^2\sin^2(\theta)}} \\
&= \int \frac{a\cos(\theta)d\theta}{a\sqrt{1-\sin^2(\theta)}} \\
&= \int \frac{\cos(\theta)d\theta}{\cos(\theta)} \\
&= \theta + c \\
&= \arcsin\left(\frac{x}{a}\right) + c.
\end{aligned}$$

Question

Find $\int \cos(x)\sin(x)dx$.

Solution

Use $u = \sin(x)$ so that $\frac{du}{dx} = \cos(x)$. Then

$$\begin{aligned}\int \cos(x)\sin(x)dx &= \int u du \\ &= \frac{u^2}{2} + c \\ &= \frac{\sin^2(x)}{2} + c.\end{aligned}$$

9.4 Explicit Integration Techniques

Question

Find $\int (3x+1)(3x^2+2x)^3 dx$.

Solution

Use $u = 3x^2 + 2x$ so that $\frac{du}{dx} = 6x + 2$ and $3x + 1 = \frac{1}{2}\frac{du}{dx}$. Then

$$\begin{aligned}\int (3x+1)(3x^2+2x)^3 dx &= \int \frac{u^3}{2} du \\ &= \frac{u^4}{8} + c \\ &= \frac{1}{8}(3x^2+2x)^4 + c.\end{aligned}$$

9.4 Explicit Integration Techniques

Our next integral technique is essentially the opposite of the product rule.

Integration by Parts

Recall that if $u(x)$ and $v(x)$ are two differentiable functions, then $(uv)' = u'v + uv'$. Therefore

$$\int uv'dx = uv - \int u'vdx.$$

This is useful if, for whatever reason, $\int u'v$ happens to be easier to solve than $\int uv'$.

Question

Find $\int x^3 \ln(x) dx$.

Solution

Use $u = \ln(x)$ and $v = \frac{x^4}{4}$, so that $u' = \frac{1}{x}$ and $v' = x^3$. Then

$$\begin{aligned}\int x^3 \ln(x) &= \frac{x^4}{4}\ln(x) - \int \frac{x^4}{4} \cdot \frac{1}{x} dx \\ &= \frac{x^4}{4}\ln(x) - \frac{1}{4}\int x^3 dx dx \\ &= \frac{x^4}{4}\ln(x) - \frac{x^4}{16} + c\end{aligned}$$

Question

Find $\int xe^x dx$.

Solution

Use $u = x$ and $v = e^x$, so that $u' = 1$ and $v' = e^x$. Then

$$\begin{aligned}\int xe^x dx &= xe^x - \int e^x \\ &= (x-1)e^x + c.\end{aligned}$$

9.4 Explicit Integration Techniques

What is $\int \frac{1}{x} dx$? We know that $\ln(x)' = \frac{1}{x}$, but the logarithm function is only defined for positive x. Therefore, we are allowed to say that

$$\int \frac{1}{x} = \ln(x),$$

but only for positive x. On the other hand, if $x < 0$, then by the chain rule, $\ln(-x)' = \frac{-1}{-x} = \frac{1}{x}$. Therefore, the general antiderivative for $\frac{1}{x}$ is given by

$$\int \frac{1}{x} = \ln(|x|) + c.$$

Our next integral technique is useful for integrating rational functions, that is, ratios of polynomials.

Integration by Partial Fractions

Suppose that $f(x) = \frac{P(x)}{Q(x)}$, where P and Q are polynomials. To integrate, you should factorise Q into powers of first or second order polynomials. Then you should express $\frac{P(x)}{Q(x)}$ as a sum of fractions, where the denominators are the factors of Q found earlier.

For example, if you are required to integrate $\frac{cx+d}{(x-a)(x-b)}$, you should find A and B so that

$$\frac{cx+d}{(x-a)(x-b)} = \frac{A}{x-a} + \frac{B}{x-b},$$

and integrate that instead.

Question

Find $\int \frac{(x+2)dx}{x^2+x}$.

Solution

The denominator factors as $x(x+1)$, so we aim to find A, B so that

$$\frac{A}{x} + \frac{B}{x+1} = \frac{x+2}{x^2+x}.$$

Then $\frac{A(x+1)}{x(x+1)} + \frac{Bx}{(x+1)x} = \frac{x+2}{x^2+x}$ so that $A = 2$ and $A + B = 1$, so $B = -1$. Therefore

$$\int \frac{(x+2)dx}{x^2+x} = \int \frac{2dx}{x} - \int \frac{dx}{x+1} = 2\ln|x| - \ln|x+1| + C.$$

Question

Find $\int \frac{dx}{x^2-a^2}$ for $a \neq 0$.

Solution

The denominator factors as $(x-a)(x+a)$, so we aim to find A, B so that

$$\frac{1}{x^2-a^2} = \frac{A}{x+a} + \frac{B}{x-a} = \frac{(A+B)x - a(A-B)}{x^2-a^2}.$$

Then $A+B=0$ and $A-B=-\frac{1}{a}$; solving gives $A=-\frac{1}{2a}$ and $B=\frac{1}{2a}$. Therefore

$$\begin{aligned}\int \frac{dx}{x^2-a^2} &= -\int \frac{dx}{2a(x+a)} + \int \frac{dx}{2a(x-a)} \\ &= \frac{1}{2a}\left(\ln|x-a| - \ln|x+a|\right) + C\end{aligned}$$

9.5 Volume Integrals

Recall that the area under the curve $y = f(x)$ between is found with

$$\int_a^b f(x)dx = \lim_{n\to\infty} \sum f(x_i^*)(x_i - x_{i-1})$$

where n describes the maximal length of some partition of $[a, b]$.

What about finding volume under a surface $z = f(x, y)$, above a region R?

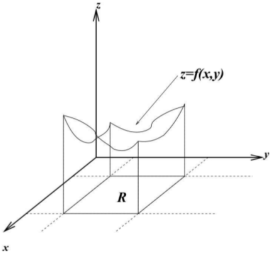


Figure 59: What is the volume V under the surface?

9.5 Volume Integrals

If our region R is the rectangle $R = \{(x, y) \in \mathbb{R}^2 | a \leq x \leq b, c \leq x \leq d\}$, we can approximate the volume by dividing $[a, b]$ into m subintervals of length $\Delta x = \frac{b-a}{m}$, and $[c, d]$ into n subintervals of length $\Delta y = \frac{d-c}{n}$.

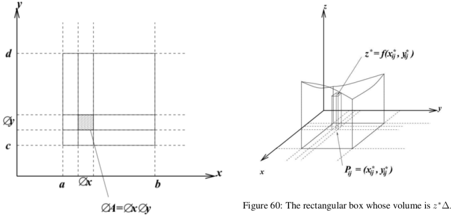


Figure 60: The rectangular box whose volume is $z^*\Delta A$.

Then the volume can be approximated by picking a point (x_{ij}^*, y_{ij}^*) in each smaller rectangle $[x_{i-1}, x_i] \times [y_{i-1}, y_i]$ and having

$$V \approx \sum_{i,j} f(x_{ij}^*, y_{ij}^*)\Delta x \Delta y.$$

Similar ideas can be used for non-rectangular regions.

9.5 Volume Integrals

Definition (Volume Integrals)

Let f be a continuous function in the region $R \subset \mathbb{R}^2$. Then the associated volume is

$$V = \lim_{m\to\infty} \lim_{n\to\infty} \sum_{i,j} f(x_{ij}^*, y_{ij}^*)\Delta A.$$

Whenever this limit exists, it is denoted $\int \int_R f(x,y)dA$.

Theorem (Properties of Volume Integrals)

- $\int \int_R (f \pm g)dA = \int \int_R fdA + \int \int_R gdA$;
- $\int \int cfdA = c \int \int fdA$;
- If R_1, R_2 is a partition of R, then $\int \int_R fdA = \int \int_{R_1} fdA + \int \int_{R_2} fdA$;
- If $f \geq g$ in R, then $\int \int_R fdA \geq \int \int_R gdA$.

9.5 Volume Integrals

The idea of *iterated integrals* gives us a concrete way to actually evaluate volume integrals using single-variable integrals we are already familiar with.
Indeed, if $R = [a, b] \times [c, d] \in \mathbb{R}^2$ is a rectangle, then

$$\int\int_R f(x, y)dA = \int_a^b \left(\int_c^d f(x, y)dy \right) dx = \int_c^d \left(\int_a^b f(x, y)dx \right) dy,$$

where $\int_c^d f(x, y)dy$ means integrating f with respect to y (keeping x fixed), and $\int_a^b f(x, y)dx$ means integrate f with respect to x (keeping y fixed).

9.5 Volume Integrals

Question

Evaluate $\int_0^2 \int_1^3 x^2 ydydx$.

Solution

The inner integral is $\int_1^3 x^2 ydy$; for a fixed x, this evaluates to $x^2 \left(\frac{y^2}{2}\right)|_1^3 = 4x^2$. Therefore

$$\begin{aligned}\int_0^2 \int_1^3 x^2 ydydx &= \int_0^2 4x^2 dx \\ &= \frac{4x^3}{3}|_0^2 \\ &= \frac{32}{3}.\end{aligned}$$

9.5 Volume Integrals

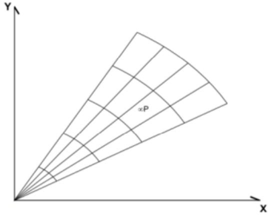
Suppose we have a function f to integrate over a non-rectangular domain D. Then we can integrate it by considering a rectangle R which surrounds D, and integrate the function

$$F(x, y) = \begin{cases} f(x, y) \text{ if } (x, y) \in D \\ 0 \text{ if } (x, y) \in R \setminus D \end{cases}$$

instead, using the sub-rectangle partition. This will converge provided f is continuous on D, and the boundary of D is 'good enough'.

9.5 Volume Integrals

If the function f and the domain D have some sort of circular symmetry, it is convenient to integrate using *polar co-ordinates*. This involves using $x = r\cos(\theta)$, $y = r\sin(\theta)$ and breaking up our domain accordingly:



.

The area of a component containing the point $P = (r^*, \theta^*)$ is approximately $r^*\Delta r\Delta\theta$. So the integral over this patch is approximately $f(r^*, \theta^*)r^*\Delta r\Delta\theta$. Therefore the volume integral is

$$\int\int_D f(r\cos(\theta), r\cos(\theta))rdrd\theta.$$

9.5 Volume Integrals

Question

Evaluate $\int\int_D e^{-x^2-y^2} dxdy$, where D is the region bounded by the circle $x^2+y^2=R^2$.

Solution

In polar co-ordinates, the region is $\{(r,\theta)|0\leq r\leq R, 0\leq\theta\leq 2\pi\}$. Therefore

$$\begin{aligned}\int\int_D e^{-x^2-y^2} dxdy &= \int_0^{2\pi}\left(\int_0^R e^{-r^2} rdr\right) d\theta \\ &= 2\pi\left(\frac{-e^{-r^2}}{2}\right)_0^R \\ &= \pi(1-e^{-R^2}).\end{aligned}$$

9.5 Volume Integrals

Question

Evaluate $\int\int_{\mathbb{R}^2} e^{-x^2-y^2} dxdy$.

Solution

Using the polar co-ordinates, this is $0 \leq \theta \leq 2\pi$ and $0 \leq r < \infty$, so like before, we find

$$\int\int_{\mathbb{R}^2} e^{-x^2-y^2} dxdy = \int_0^{2\pi}\int_0^{\infty} e^{-r^2} r dr d\theta$$
$$= 2\pi \left(\frac{-e^{-r^2}}{2}\right)_0^{\infty} = \pi.$$

9.6 Improper Integrals

Improper integrals just mean integrals over unbounded regions.

Definition (Integrals over semi-infinite domains)

Suppose $\int_a^t f(x)dx$ exists for all $t \geq a$. Then $\int_a^\infty f(x)dx = \lim_{t\to\infty} \int_a^t f(x)dx$. Similarly, if $\int_t^b f(x)dx$ exists for all $t \leq b$, then $\int_{-\infty}^b f(x)dx = \lim_{t\to-\infty} \int_t^b f(x)dx$. These integrals are said to be *convergent* if these limits exist, and *divergent* otherwise.

Definition (Integrals over all real numbers)

If for each $a \in \mathbb{R}$, the integrals $\int_{-\infty}^a f(x)dx$ and $\int_a^\infty f(x)dx$ both exist, then we say

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{a} f(x)dx + \int_{a}^{\infty} f(x)dx.$$

It can be shown that this definition does not depend on a.