第三周　　随机变量及其概率分布

3.1、随机变量及分布函数

函数的概念可以推广到自变量不是实数的情形。如：两点间的距离可作为以一对点为自变量的函数；三角形的周长为定义在三角形集合上的函数。随机变量是一个从样本空间Ω到实数集合R的函数。

考虑一个包含n个人样本空间 $\Omega=\{\omega_1,\omega_2,\cdots,\omega_n\}$

记$A(\omega)$表示样本ω的年龄，$\omega\to A(\omega)$，得到一个$\Omega\to Z^+$ 映射。

类似的我们可以用映射关系

$\omega\to H(\omega)$，$\omega\to W(\omega)$，$\omega\to I(\omega)$ 表示身高、体重、收入等

　　对于我们所关心的随机现象，其样本点的形式多种多样，除了公司的员工，还可能是甲乙比赛的胜负结果，可能是抽取到的几个白球、黑球等等。这些表现形式是无法进行数学计算的，只有将它们对应为实数，才能进行进一步的定量的处理。引入随机变量的概念就是出于这个目的，它将随机试验的结果数量化，从而可以用简洁的数学语言描述繁杂的随机问题，同时提高处理随机问题的效率。随机变量的引入可以说是概率论发展中的一个重要的里程碑。下面看几个例子。

****

例 3.1.1 随机将一枚硬币投掷两次，投出正面记为“正”，反面记为“反”，

则此随机试验的样本空间为 $\Omega=\{$正正，正反，反正，反反$\}$。

用X表示正面的次数，则得到样本空间到实数集合的映射，$X:\Omega\to R$，

函数取值为：$X($正正$)=2$，$X($正反$)=1$，$X($反正$)=1$，　$X($反反$)=0$。

事件“两次投掷中，恰有 1 次正面”$=\{\omega:\ X(\omega)=1\}$，也可简写为$\{X=1\}$。

$P(\{X=k\})$可简写为$P(X=k)$，$k=0,1,2$。

若假设硬币是均匀的，则Ω中的所有样本点是等可能出现的，因此，

$P(X=0)=\frac{1}{4}$，　$P(X=1)=\frac{1}{2}$，　$P(X=2)=\frac{1}{4}$。

****

例 3.1.2 随机将一枚均匀的硬币投掷 3 次，则样本空间为

$\Omega=\{$正正正，正正反，正反正，正反反，反正正，反正反，反反正，反反反$\}$

(1) 如果需要区分每一个基本结果，则可定义随机变量，

ω :	正正正	正正反	正反正	正反反	反正正	反正反	反反正	反反反
$X(\omega)$:	0	1	2	3	4	5	6	7

(2) 如果关心的是得到正面的次数，则可定义随机变量

ω :	正正正	正正反	正反正	正反反	反正正	反正反	反反正	反反反
$X(\omega)$:	3	2	2	1	2	1	1	0

(3) 若制定一个赌博规则，三次都得到正面收益 100 元，其余结果均输掉 10 元，则可定义随机变量

ω :	正正正	正正反	正反正	正反反	反正正	反正反	反反正	反反反
$X(\omega)$:	100	−10	−10	−10	−10	−10	−10	−10

****

随机变量的定义

设随机试验的样本空间为Ω，如果对于每一个$\omega\in\Omega$，有一个实数$X(\omega)$与之对应，这样得到一个定义在Ω上的单值函数$X=X(\omega)$，称$X(\omega)$为随机变量，简记为X。

也就是说，随机变量X是一个从Ω到实数域R的函数，它的定义域为Ω，它的值域$X(\Omega)$为R或R的一个子集。通常用英文的大写字母X,Y,Z等来表示随机变量，其取值用小写字母x,y,z等表示。

****

事件的示性函数

函数 $I_A(\omega)=\begin{cases}1 & \omega\in A\\ 0 & \omega\notin A\end{cases}$，称为事件$A$的示性函数或标志函数。

$I_A(\omega)$即为随机变量。

这个例子表明随机事件是随机变量一个特例。随机变量是随机事件的推广，它拓展了概率论的研究范围与手段。今后我们的研究对象主要集中在随机变量。

****

随机变量的分布函数

设 X 是一个随机变量，对任意实数 x，

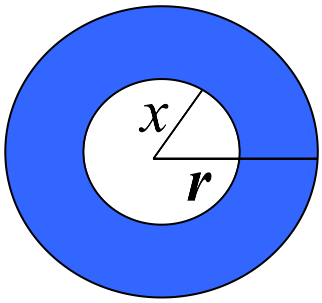
定义 $F(x)=P(X\le x)$ 为随机变量 X 的分布函数，

且称 X 服从 $F(x)$，记为 $X\sim F(x)$，有时也记作 $F_X(x)$。

$\lim\limits_{x\to-\infty}F(x)=0$，　　$\lim\limits_{x\to+\infty}F(x)=1$。

****

例 3.1.3 在半径为 r 的圆内随机抛一点，假设这个点落在圆内任何位置是等可能的，求此点到圆心距离 X 的分布函数 $F(x)$。



解：　$F(x)=P(X\le x)=\begin{cases}0 & x<0\\ \left(\dfrac{x}{r}\right)^2 & 0\le x<r\\ 1 & x\ge r\end{cases}$

****

3.2 离散型与连续型随机变量

在处理实际概率问题的时候，为了便于使用，人们按照随机变量所具有的不同共性特征，将随机变量进行分类，针对不同随机变量的特点，总结出更适合分析与计算的表达形式。最经常使用的是离散型与连续型这两种随机变量。

****

离散型随机变量

如果随机变量 X 所有可能的取值是有限或可列多个，则称为离散型随机变量，

则其分布可表示为

$$\begin{array}{c|ccccc} X & x_1 & x_2 & \cdots & x_n & \cdots \\ \hline P & p_1 & p_2 & \cdots & p_n & \cdots \end{array} \quad 或 \quad \begin{pmatrix} x_1 & x_2 & \cdots & x_n & \cdots \\ p_1 & p_2 & \cdots & p_n & \cdots \end{pmatrix}$$

这种表示称为分布列。

其中 $p_i = P(X = x_i) \geq 0 \quad (i = 1,2,\cdots n,\cdots)$，$\sum_{i=1}^{\infty} p_i = 1$，$F(x) = \sum_{x_i \leq x} p_i$。

分布函数图形为阶梯函数。

这里补充说明一下什么是“可列”，可列也称“可数”，是指某种无穷集合元素的个数。如果一个包含无穷个元素的集合，它的所有元素都可以用自然数进行编号，用 a1,a2,a3 的方式完全罗列出来，则称这个集合的元素个数是可列多个或可数多个。例如，所有正的偶数构成的集合，将集合元素按照从小到大的顺序排列，则其第 n 个元素即为 2n，一直罗列下去，就可以表达出所有元素。还可以证明有理数集合的元素也是可数的，即可以设定一个规则，给每一个有理数赋予一个特定编号，一个一个数下去，就能够数完全体有理数。但并不是所有无穷集合都是可数的，实数集合，即使是限定在(0,1）区间上，其元素的个数也是不可数的，无法给出一种编号规则，遍历该集合所有的元素。这些证明无法在这里展开。希望同学们对可列或可数的概念有一个感性的了解就好了。19 世纪末，德国数学家康托尔开创了集合论，并引入“可数”这个概念。康托尔开创的理论，是自古希腊时代的二千多年以来，人类认识史上第一次给无穷建立起抽象的形式符号系统和确定的运算，它从本质上揭示了无穷的特性。这些理论的应用渗透到数学的各个分支，成为实变函数论、代数拓扑、群论和泛函分析等等理论的基础。

****

例 3.2.1 两点分布：$\begin{pmatrix} a_0 & a_1 \\ p & q \end{pmatrix}$，其中 $p + q = 1$，$p,q \geq 0$，

最常用的两点分布是 0-1 分布（$a_0 = 0, a_1 = 1$ 的情形，也称伯努利分布）

分布律：$\begin{pmatrix} 0 & 1 \\ p & q \end{pmatrix}$，分布函数：$F(x) = \begin{cases} 0 & x < 0 \\ p & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$

****

连续型随机变量

设随机变量 X 的分布函数为 $F(x)$，如果存在非负可积函数 $f(x)$ $(x\in R)$，使得 $\forall x\in R$，有 $F(x)=\int_{-\infty}^{x} f(t)dt$，则称 X 为连续型随机变量，$f(x)$ 称为 X 的概率密度函数，简称密度函数（probability density function，常缩写为 pdf）。

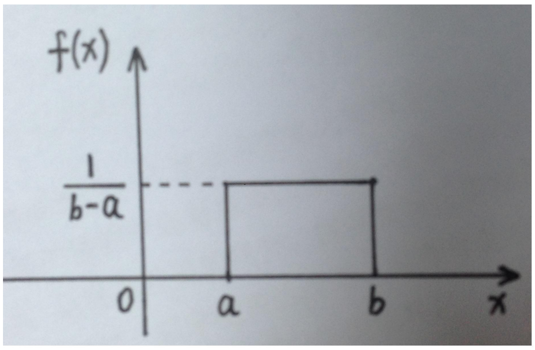
概率密度函数一定满足：$\int_{-\infty}^{+\infty} f(x)dx=1$。

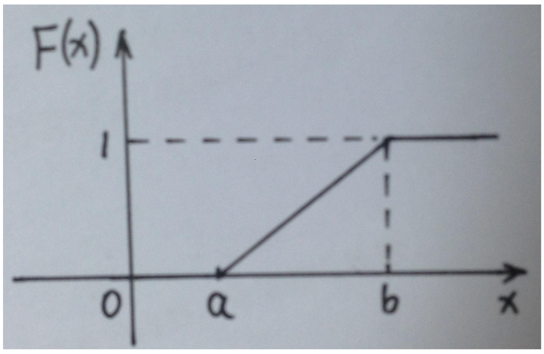
连续型随机变量的分布函数一定是 R 上的连续函数，但分布函数在 R 上连续的随机变量不一定都是连续型的（后面介绍的 Cantor 分布就是一个例子）。

****

例 3.2.2 均匀分布 $-\infty<a<b<+\infty$

$$f(x)=\begin{cases}\dfrac{1}{b-a} & x\in[a,b] \\ 0 & 其他\end{cases}, \qquad F(x)=\begin{cases}0 & x<a \\ \dfrac{x-a}{b-a} & a\le x<b \\ 1 & x\ge b\end{cases}$$





****

例 3.2.3 柯西分布 $f(x)=\dfrac{1}{\pi}\dfrac{1}{1+x^2}$， $-\infty<x<+\infty$

$$\int_{-\infty}^{+\infty} f(x)dx=\int_{-\infty}^{+\infty}\frac{1}{\pi}\frac{1}{1+x^2}dx=\frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{1}{1+x^2}dx$$

$$=\frac{1}{\pi}\arctan x\Big|_{-\infty}^{+\infty}=\frac{1}{\pi}\left[\frac{\pi}{2}-\left(-\frac{\pi}{2}\right)\right]=1$$

****

例 3.2.4 某电子元件的寿命 X（单位：小时）服从以下分布

$$f(x)=\begin{cases}\dfrac{1000}{x^2} & X>1000\\ 0 & \text{其他}\end{cases}。$$

（1） 任取一只，其寿命不超过 1500 小时的概率？

（2） 已知一只元件已经使用 1500 小时，求该元件能够继续使用 500 小时的概率？

解：（1） $P(X\le 1500)=\int_{-\infty}^{1500}f(x)dx=\int_{1000}^{1500}\dfrac{1000}{x^2}dx=-\dfrac{1000}{x}\Big|_{1000}^{1500}=\dfrac{1}{3}$

（2） $P(X\ge 2000|X\ge 1500)=\dfrac{P(X\ge 2000,X\ge 1500)}{P(X\ge 1500)}=\dfrac{P(X\ge 2000)}{1-P(X\le 1500)}$

$$=\frac{\int_{2000}^{+\infty}\frac{1000}{x^2}dx}{2/3}=\frac{1/2}{2/3}=\frac{3}{4}。$$

****

3.3 分布函数的性质与特殊的例子

分布函数的性质

任意随机变量的分布函数 $F(x)$ 都具有如下三条基本性质

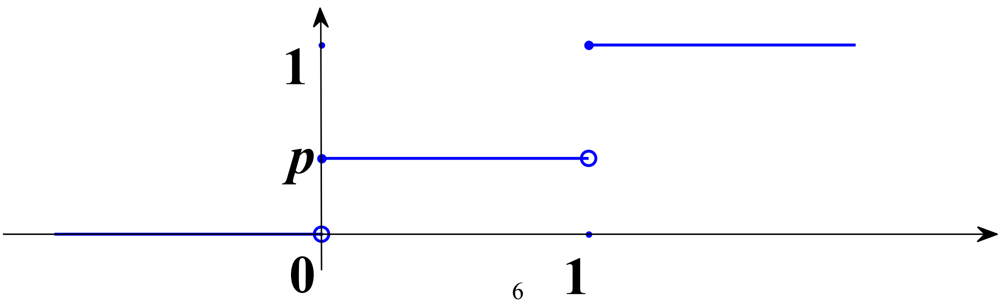
（1） 单调性　　$F(x)$ 是 $(-\infty,+\infty)$ 上单调非减函数，即 $\forall x_1<x_2, F(x_1)\le F(x_2)$；

（2） 有界性　　$\forall x\in R$，有 $0\le F(x)\le 1$，

$$F(-\infty)=\lim_{x\to-\infty}F(x)=0,\quad F(\infty)=\lim_{x\to\infty}F(x)=1;$$

（3） 右连续性　$F_X(x)$ 是关于 x 的右连续函数，即 $\forall x_0\in R$，有

$$\lim_{x\to x_0^+}F(x)=F(x_0),\quad F(x_0+0)=F(x_0)$$



****

利用分布函数计算随机变量在不同区域上的概率

$$P(a<X\le b)=P(X\le b)-P(X\le a)=F(b)-F(a)$$

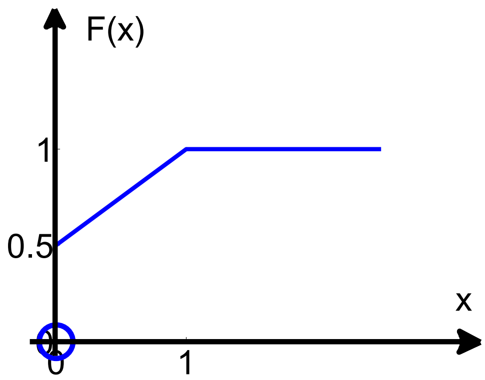
$$P(X=a)=P(X\le a)-P(X<a)=F(a)-\lim_{x\to a^-}F(x)=F(a)-F(a-0)$$

$$P(X\ge b)=1-P(X<b)=1-[P(X\le b)-P(X=b)]=1-F(b-0),$$

$$P(a\le X\le b)=P(a<X\le b)+P(X=a)=F(b)-F(a)+P(X=a)=F(b)-F(a-0)$$

****

既非离散型也非连续型的随机变量



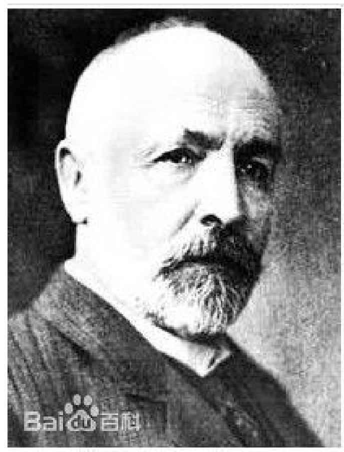
例 3.3.1 $F(x)=\begin{cases}0 & x<0\\ \dfrac{x+1}{2} & 0\le x<1\\ 1 & x\ge 1\end{cases}$

$$F(x)=\frac{1}{2}\left(F_1(x)+F_2(x)\right)$$

$$F_1(x)=\begin{cases}0 & x<0\\ 1 & x\ge 0\end{cases},\quad F_2(x)=\begin{cases}0 & x<0\\ x & 0\le x<1\\ 1 & x\ge 1\end{cases}$$

****

例 3.3.2 既非连续也非离散的分布，康托尔（Cantor）分布

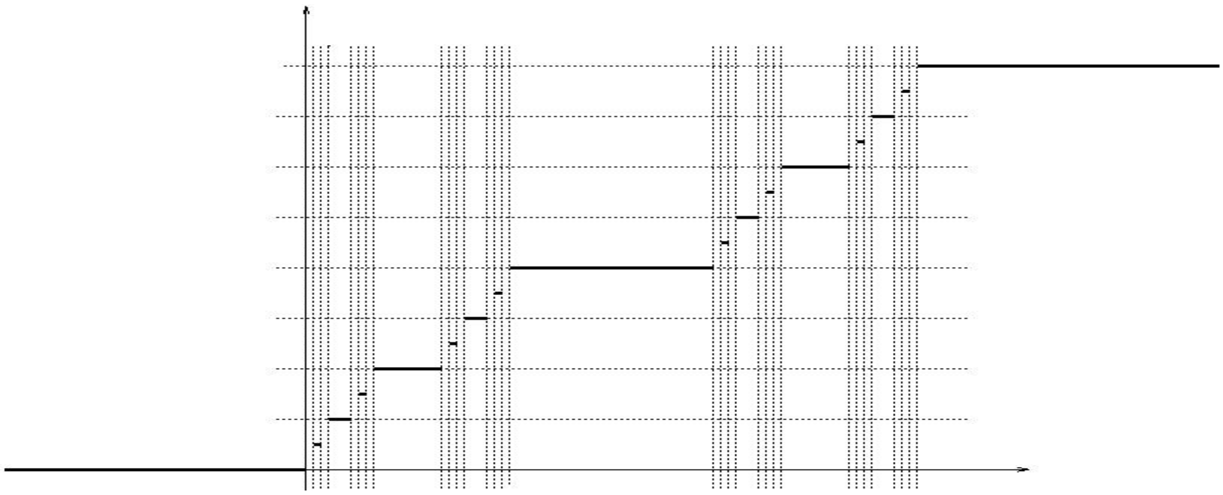


$$C_1=\left(\frac{1}{3},\frac{2}{3}\right),\quad C_2=\left(\frac{1}{9},\frac{2}{9}\right)\cup\left(\frac{7}{9},\frac{8}{9}\right),$$

$$C_3=\left(\frac{1}{27},\frac{2}{27}\right)\cup\left(\frac{7}{27},\frac{8}{27}\right)\cup\left(\frac{19}{27},\frac{20}{27}\right)\cup\left(\frac{25}{27},\frac{26}{27}\right),$$

$$C_4=\cdots$$

定义函数 $F(x)=\begin{cases}0 & x\le 0\\ \dfrac{2k-1}{2^n} & 当x\in C_n的第k个子集时，0<k\le 2^{n-1}, n=1,2,\cdots,\\ 1 & x\ge 1\end{cases}$



对所有 $n=1,2,\cdots$，第 n 级集合 C_n 包括 2^{n-1} 个小区间，每个区间的长度为 $\dfrac{1}{3^n}$，

则所有 C_n （$n=1,2,\cdots,$）的长度和为 $\lim\limits_{n\to\infty}\left(\dfrac{1}{3}+2\cdot\dfrac{1}{3^2}+2^2\cdot\dfrac{1}{3^3}+\cdots+2^{n-1}\cdot\dfrac{1}{3^n}\right)=1$。

给 $(0,1)$ 内剩余的点赋值使 $F(x)$ 满足连续性，则 $F(x)$ 为单调不减的连续函数。

$F(x)$ 符合分布函数的定义，此函数定义了一个随机变量，记为 X。

由于在这个实数域内几乎处处有 $f(x)=F'(x)=0$，

则 $\int_{-\infty}^{\infty} f(x)dx=\int_{-\infty}^{\infty} F'(x)dx=0$，$F(x)$ 不存在密度函数，所以 X 不是连续型。

离散型随机变量：有限个或可数多个取值，其分布函数为有限或可数多个取值变化的阶梯函数，不可能在整个实数域内都是连续的。

$F(x)$ 是 R 上的连续函数，所以 X 也不是离散型。

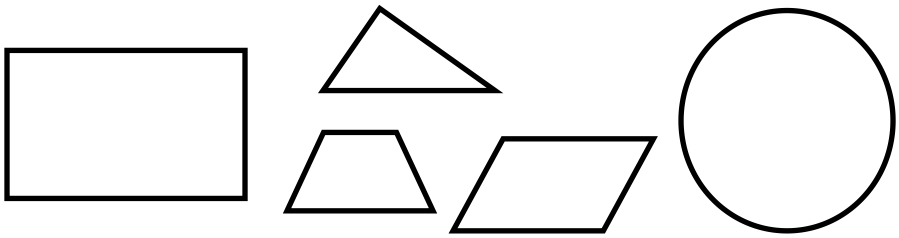
所以由 $F(x)$ 定义的随机变量既非离散型也非连续型。

****

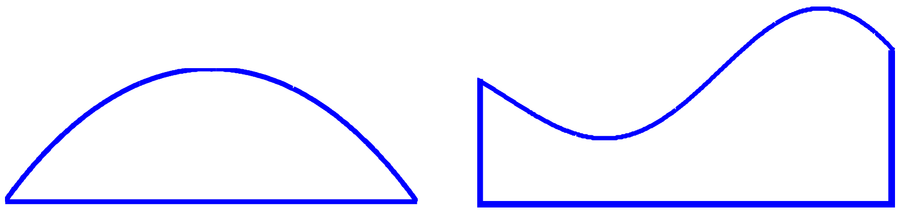
3.4 概率论所需微积分要点回顾

面积的计算

简单



困难



****

$$\text{面积} = \lim_{\max\{\Delta t_k\}\to 0} \sum_{k=1}^{n} v(\xi_k)\Delta t_k = S(b) - S(a)$$

考虑曲线 $v(t)$ 围成的面积，我们不妨假设这条曲线代表某个物体的速度变化情况，

用函数 v(t) 表示，曲线的起点 a 和终点 b 分别代表两个时刻。这时，曲线在 a，b 之间围成的面积就等于在 a，b 时刻之间物体走过的路程。计算面积就可以避开永远变化着的速度函数。如果知道路程随时间变化的函数 S(t)，则总路程即为 S(b)-S(a)，v(t) 在 a，b 两点间围成的面积就等于 S(b)-S(a)，只和两个点有关，而不再依赖无穷多个点！我们还发现，S(t) 和 v(t) 有内在的联系，路程的变化率等于速度，所以 S(t) 的导数等于 v(t)。要计算 v(t) 围成的面积，我们不用考虑它在无穷个点上的变化，只要找到哪个函数的导数是 v(t)，则这个面积用那个函数两个点的取值的差即可得到。

也就是无穷多个点所围面积实际上被另一个函数在两个点的信息所控制。如果我们得到这个控制函数，无穷多个点的变化就都化解了，而归结为两个点的信息。

我们会的总是简单的，所谓规律往往就是操纵变化过程的简单规则，关键是我们能否认识到复杂表象下是否有简单的规则在背后操纵它们。如果发现了，我们的认识就提高了一步。当然，所谓简单也不是绝对的，积累的理论和实践经验越丰富，此

时简单的内涵就越丰富。可以说理解清晰了，就是简单了。

****

微积分基本定理：$\int_a^b f(x)dx = F(b) - F(a)$，其中 $F'(x) = f(x)$。

常用函数的导数公式　　常用函数的积分公式

复合函数的求导　　积分的换元法

两个函数乘积的导数公式　　分部积分法

****

基本导数公式

1. $(C)' = 0$

2. $(x^a)' = ax^{a-1}$

3. $(\sin x)' = \cos x$

4. $(\cos x)' = -\sin x$

5. $(e^x)' = e^x$

6. $(\ln x)' = \dfrac{1}{x}$

7. $(\arcsin x)' = \dfrac{1}{\sqrt{1-x^2}}$

8. $(\arctan x)' = \dfrac{1}{1+x^2}$

基本积分公式

1. $\int_a^b 1dx = x\Big|_a^b = b - a$

2. $\int_a^b x^\alpha dx = \dfrac{x^{\alpha+1}}{\alpha+1}\Bigg|_a^b = \dfrac{b^{\alpha+1} - a^{\alpha+1}}{\alpha+1}$

3. $\int_a^b \cos xdx = \sin x\Big|_a^b = \sin b - \sin a$

4. $\int_a^b \sin xdx = -\cos x\Big|_a^b = \cos a - \cos b$

5. $\int_a^b e^x dx = e^x\Big|_a^b = e^b - e^a$

6. $\int_a^b \dfrac{1}{x}dx = \ln x\Big|_a^b = \ln\dfrac{b}{a}\ \ (0 < a < b)$

7. $\int_a^b \dfrac{1}{\sqrt{1-x^2}}dx = \arcsin x\Big|_a^b = \arcsin b - \arcsin a\ \ \ (-1 \le a \le b \le 1)$

8. $\int_a^b \frac{1}{1+x^2}dx = \arctan x\Big|_a^b = \arctan b - \arctan a$

****

复合函数的求导公式

$$\frac{df\left(g(x)\right)}{dx} = \frac{df\left(g(x)\right)}{dg(x)} \cdot \frac{dg(x)}{dx} = f'\left(g(x)\right)g'(x)$$

$$\Rightarrow df\left(g(x)\right) = f'\left(g(x)\right)g'(x)dx$$

$$\Rightarrow \int_a^b df\left(g(x)\right) = \int_a^b f'\left(g(x)\right)g'(x)dx$$

换元积分公式

$$\int_a^b f'\left(g(x)\right)g'(x)dx = \int_a^b df\left(g(x)\right) = f\left(g(x)\right)\Big|_a^b = f\left(g(b)\right) - f\left(g(a)\right)$$

$$\int_a^b f'\left(g(x)\right)g'(x)dx = \int_a^b f'\left(g(x)\right)dg(x) \xrightarrow[=]{\text{令 } y=g(x)} \int_{g(a)}^{g(b)} f'(y)dy$$

****

例 3.4.1 计算积分$\int_a^b xe^{-x^2}dx$

$$\int_a^b xe^{-x^2}dx = -\frac{1}{2}\cdot\int_a^b e^{-x^2}d\left(-x^2\right) = -\frac{1}{2}\cdot\int_a^b de^{-x^2} = -\frac{1}{2}e^{-x^2}\Big|_a^b = \frac{e^{-a^2}-e^{-b^2}}{2}$$

$$\int_a^b xe^{-x^2}dx = -\frac{1}{2}\cdot\int_a^b e^{-x^2}d\left(-x^2\right) \xrightarrow[=]{\text{令 } y=-x^2} -\frac{1}{2}\cdot\int_{-a^2}^{-b^2} e^y dy = -\frac{1}{2}e^y\Big|_{-a^2}^{-b^2} = \frac{e^{-a^2}-e^{-b^2}}{2}$$

****

乘积函数的求导公式

$$\left[f(x)g(x)\right]' = f'(x)g(x) + f(x)g'(x)$$

$$\xrightarrow{\text{两边积分}} \int_a^b \left[f(x)g(x)\right]' dx = \int_a^b f'(x)g(x)dx + \int_a^b f(x)g'(x)dx$$

$$\Rightarrow \quad f(x)g(x)\Big|_a^b = \int_a^b g(x)df(x) + \int_a^b f(x)dg(x)$$

分部积分公式 $\int_a^b f(x)dg(x) = f(x)g(x)\Big|_a^b - \int_a^b g(x)df(x)$

****

例 3.4.2 计算积分$\int_a^b xe^{-x}dx$

$$\int_a^b xe^{-x}dx = \int_a^b xd\left(-e^{-x}\right) = -xe^{-x}\Big|_a^b - \int_a^b \left(-e^{-x}\right)dx$$

$$= -xe^{-x}\Big|_a^b - e^{-x}\Big|_a^b$$

$$= ae^{-a} - be^{-b} - e^{-b} + e^{-a} = (a+1)e^{-a} - (b+1)e^{-b}$$