

我是8位的

I am 8 bits, what about you?

随笔 - 205, 文章 - 0, 评论 - 103, 阅读 - 101万

导航

博客园
首页
新随笔
联系
订阅
管理

2022年3月						
日	一	二	三	四	五	六
27	28	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2
3	4	5	6	7	8	9

公告

你的支持是我的动力
欢迎关注微信公众号“我是8位的”



昵称：我是8位的
园龄：4年7个月
粉丝：288
关注：5
+加关注

盖楼抽奖

#她的梦想在发光#

HWD科技女性故事有奖征集

分享最打动你的科技女性故事

活动时间：2022年3月8日-3月18日

马上参与



搜索

找找看

谷歌搜索

常用链接

我的随笔
我的评论
我的参与
最新评论
我的标签

积分与排名

积分 - 457097
排名 - 1198

概率笔记5——概率分布

分布函数（英文Cumulative Distribution Function, 简称CDF），是概率统计中重要的函数，正是通过它，可用数学分析的方法来研究随机变量。分布函数是随机变量最重要的概率特征，分布函数可以完整地描述随机变量的统计规律，并且决定随机变量的一切其他概率特征。

从事件到函数

我们已经很清楚函数的概念， $g = g(x)$ 是一个典型的函数，输入数据经过 $g(x)$ 的处理后得到了一个新的输出 g 。在概率当中，也存在类似的定义。

例如一个样本空间有一系列随机事件 $\Omega = \{\omega_1, \omega_1, \omega_1 \dots \omega_n\}$ ，那么将存在一个函数，这个函数把事件映射为一个实数：

$$X = X(\omega_i)$$

这样做是为了用数学去表达事件——函数最终将转换为数，有了数，我们就能利用很多已知的工具去处理概率问题。如果 Ω 表示球队的比赛事件，那么 $\Omega = \{\text{胜}, \text{负}, \text{平}\}$ ，一个典型的 X 转换就是：胜 $\rightarrow 1$ ，负 $\rightarrow -1$ ，平 $\rightarrow 0$ 。二进制的0和1也能表达很多诸如开/关、升/降等事件。

分布函数

有了函数 X ，就可以进而将事件的概率转换为普通的函数，于是有了分布函数的定义：

$$F(x) = P\{X \leq x\}$$

$F(x)$ 就是分布函数，它表示 $X \leq x$ 的概率。举例来说，如果一个人的身高是1.75m，这个人的身高在全国的分布就是所有小于等于1.75m的人在全国的比例。看起来英文Cumulative Distribution Function更容易理解， $F(x)$ 就是概率的积累。

需要注意的是，此处的大 X 和小 x 都是一个具体的实数，小 x 的取值范围是 $-\infty \leq x \leq +\infty$ ，这是一个什么梗？

这要从坐标系说起了。

随笔分类 (211)

- ★★资源下载★★(1)
- Java并发编程(1)
- 程序员的数学(24)
- 单变量微积分(31)
- 多变量微积分(24)
- 概率(24)
- 机器学习(27)
- 软件设计(1)
- 数据分析(6)
- 数据结构与算法(27)
- 随笔(5)
- 线性代数(34)
- 项目管理(2)
- 转载(4)

随笔档案 (205)

- 2021年2月(1)
- 2020年3月(2)
- 2020年2月(6)
- 2020年1月(4)
- 2019年12月(7)
- 2019年11月(15)
- 2019年9月(3)
- 2019年8月(6)
- 2019年7月(1)
- 2019年6月(8)
- 2019年5月(3)
- 2019年4月(5)
- 2019年3月(7)
- 2019年2月(3)
- 2019年1月(7)
- 更多

阅读排行榜

- 1. 使用Apriori进行关联分析 (一) (29768)
- 2. 线性代数笔记12——列空间和零空间 (28772)
- 3. FP-growth算法发现频繁项集 (一)——构建FP树(24430)
- 4. 寻找“最好” (2) ——欧拉-拉格朗日方程(23099)
- 5. 多变量微积分笔记3——二元函数的极值(22772)

评论排行榜

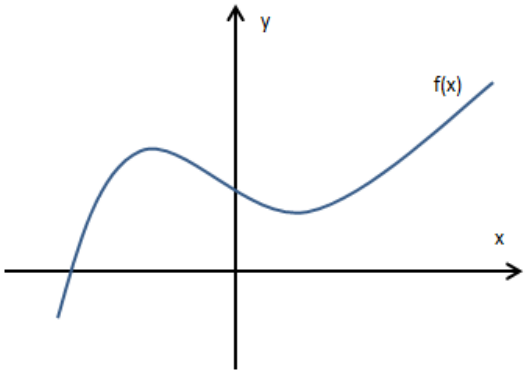
- 1. 隐马尔可夫模型 (一) (8)
- 2. 线性代数笔记12——列空间和零空间 (7)
- 3. 线性代数笔记3——向量2 (点积) (7)
- 4. FP-growth算法发现频繁项集 (一)——构建FP树(5)
- 5. 寻找“最好” (2) ——欧拉-拉格朗日方程(4)

推荐排行榜

- 1. 寻找“最好” (2) ——欧拉-拉格朗日方程(7)
- 2. FP-growth算法发现频繁项集 (一)——构建FP树(7)
- 3. 线性代数笔记3——向量2 (点积) (6)
- 4. FP-growth算法发现频繁项集 (二)——发现频繁项集(5)
- 5. 隐马尔可夫模型 (一) (5)

最新评论

- 1. Re:线性代数笔记3——向量2 (点积)
如果点积小于0, 即夹角小于90°, 这个写错了吧。应该是夹角大于90°
--猫猫猫猫大人



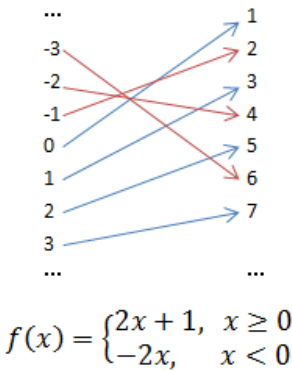
上图中的曲线是 $f(x)$, $-\infty < x < +\infty$, 对于任意的 x , 都有一个 y 能够对应。同样, 对于概率分布函数 $F(x)$ 来说, 我们也希望对任意的 x 都能找到对应的 y , 也就是 $P(X \leq x)$ 。别忘了, 我们的目的是将事件转换为数, 从而将概率转换为函数。从概率的角度来讲, $-\infty < x < +\infty$ 表示了概率的全部事件。

离散型分布

离散事件

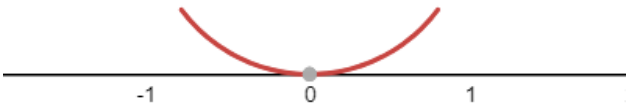
离散型事件指事件可能的取值是有限个或可列无穷个。
有限个好理解, 比如骰子的结果。可列无穷个有意思了, 它指值能够列举出来, 但是永远无法全部列举, 自然数和整数就是这样的例子。
这里有个好玩的事, 整数是无穷的, 自然数也是无穷的, 那么整数和自然数的数量哪个更多呢?

第一感觉是整数更多, 多了一倍。但真相是, 二者的数量一样多。这就需要了解数学中是怎样定义“一样多”的。在数学中, 如果两个集合能够产生一一对应的关系, 我们就可以说这两个集合的数据一样多。这个对应关系可以用一个函数表示, 比如整数和自然数的对应可以是这样:



$$f(x) = \begin{cases} 2x + 1, & x \geq 0 \\ -2x, & x < 0 \end{cases}$$

无论哪一个整数, 都能在自然数中找到唯一的对应。
整数和实数呢? 实数的个数要远远大于整数, 它们无法产生一一对应, 因为每两个实数间都有无穷多个数。这就又引出一个问题, 实数的个数与 $[-1, 1]$ 区间内的实数个数哪个多呢? 第一感觉又是实数多, 但实际上二者的个数相等。这个匪夷所思的问题可以用下图表示, 说明二者一一对应:



2. Re:线性代数笔记10——矩阵的LU分解写的很好，不过LU分解的前提是错的，LU分解只需要第三个条件，如果允许行置换就是下面写到的PLU，可以分解所有矩阵

--wiki3D

3. Re:单变量微积分笔记20——三角替换1 (sin和cos) 很nice

--尹保棕

4. Re:线性代数笔记24——微分方程和exp(At) 有些图片挂了呢

--ccchendada

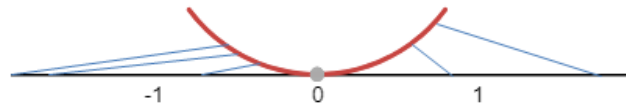
5. Re:寻找“最好” (2) ——欧拉-拉格朗日方程

提个issue，最速降线中

$v = \{2gh\}^{1/2}$ 与配图不一致，建议以起点为原点，向右伸出x轴，向下伸出y轴建立坐标系

--trustInU

上图是一个数轴，数轴上的每一个点都代表一个实数；现在把-1到1之间的线段的向上弯折，得到一个与0点相切，弧长是2的红色圆弧。现在，把数轴上的任意点与弧连线，都可以在弧上找到唯一点：



由此可见，二者的数量相等，准确的说是“势”相等。

分布函数

离散事件的每个取值都对应一个概率，它的分布率大概长成这个样子：

$$\begin{matrix} X & (x_1 & x_2 & \cdots & x_n & \cdots) \\ P & (P_1 & P_2 & \cdots & P_n & \cdots) \end{matrix}$$

它的分布函数：

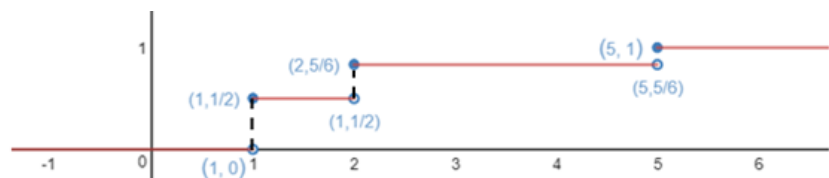
$$F(x) = P\{X \leq x\} = \sum_{x_i \leq x} P\{X = x_i\}, \quad -\infty < x < +\infty$$

在所有的分布函数中，x的取值范围都是关键，它强调了“事件”到“函数”的转换。

在射击比赛中，有大、中、小三类目标供选择，各类目标的得分和命中率如下：

	goal	big	middle	small
score		1	2	5
rate		1/2	1/3	1/6

其中score对应了x的取值，rate对应分布值F(x)，F(x)的分布曲线如下：



这里又一次强调了分布函数F(x)中x的取值是从 $-\infty$ 到 $+\infty$ 。当 $x < 1$ 时，表示没有任何目标可供射击，命中率是0； $x \leq 2$ 时，命中中型和中型以下目标的概率是 $F(2) = P(\text{middle}) + P(\text{small}) = 1/3 + 1/2 = 5/6$ ； $x \geq 5$ 时，变成了必然事件， $F(x) = 1$ 。

我们看到F(x)的取值是[0, 1]，这也是概率的取值范围；这种阶梯式的函数就是离散型随机事件的分布函数。

连续型分布

连续事件

相对于离散事件，连续事件就是随机事件是连续型的事件。这是通俗解释，看起来没错，但并不精确。

在精确定义之前先来看一个好玩的例子：一个人会在9:00~10:00到达某地，他恰巧在9:30抵达的概率是多少？

似乎很简单，但实际上不是那么回事，问题出在时间的度量上。前面说过，0~1之间的实数有无穷多个，同样，由于我们并没有指定时间的最小刻度，所以9:00~10:00之间的也有无穷多个，这相当于样本空间的事件有无穷个。如果用几何概型思考——将概率转换为长度的比例——我们会发现，9:30是时间轴上的一点，点的长度是0，所以 $P\{9:30\text{抵达}\} = 0$ 。过去一直认为0概率是不肯能发生的事件，而现在看来并不是，因为确实存在9:30抵达的可能，这有点像极限问题了，极限是0，说明无限接近0，但始终不是0。

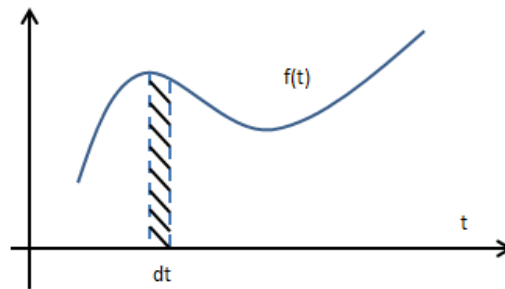
似乎出现悖论了，无数个点加在一起变成了线，点的概率又是0，那么连续事件的分布岂不是无数个0相加最终还是0？

解释前先写出连续事件的精确定义：对于某一 X ，如果存在非负可积函数 $f(x)$ ，使得

$$F(x) = P\{X \leq x\} = \int_{-\infty}^x f(t)dt, \quad -\infty < x < +\infty$$

则称 X 是连续型随机事件。

答案就是使用积分。使用 $f(t)dt$ 就可以计算微小的面积：

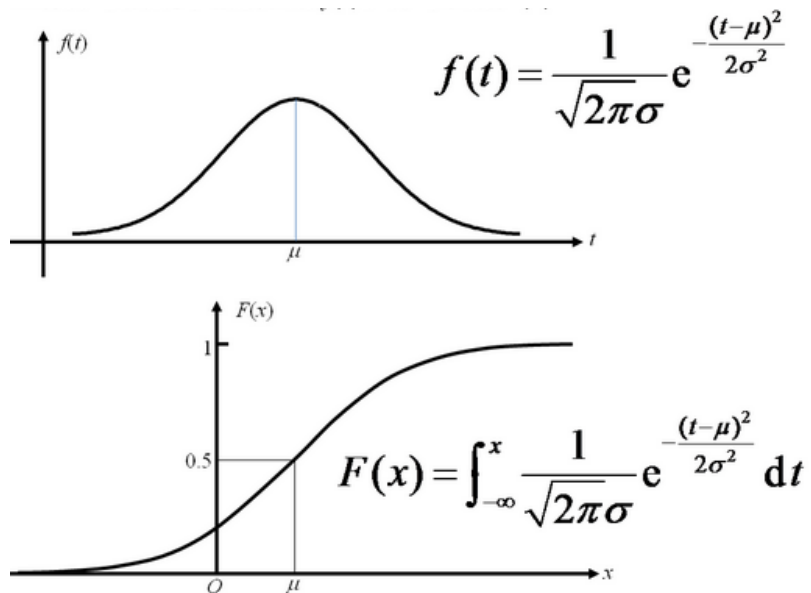


关于微分和积分的相关知识可参考：[《单变量微积分》](#)中的相关章节。

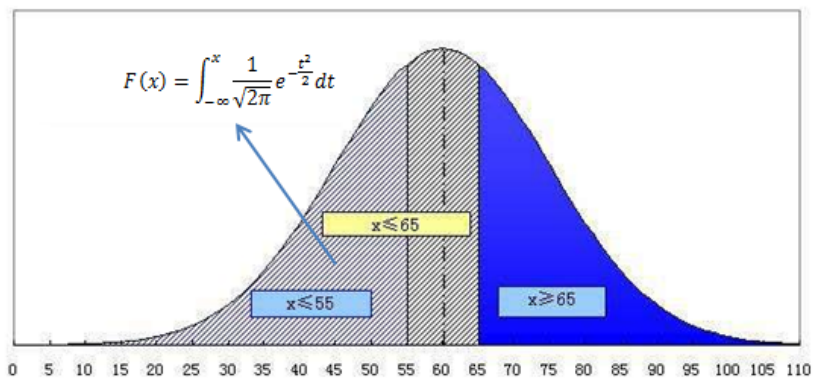
现在概率终于和积分联系在一起了，前方的视野也更加广阔起来。

分布函数

以正态分布为例：



$f(t)$ 被称为概率密度，或概率密度函数； $F(x)$ 表示 $f(t)$ 与 x 轴围成的面积：



由此可以看出，连续型随机事件的分布函数也一定是连续的。

作者：我是8位的

出处：<http://www.cnblogs.com/bigmonkey>

本文以学习、研究和分享为主，如需转载，请联系本人，标明作者和出处，非商业用途！

扫描二维码关注公众号“我是8位的”



随笔

分类：概率

标签: [分布函数](#), [概率分布](#)

好文要顶

关注我

收藏该文

我是8位的

关注 - 5

粉丝 - 288

+加关注

0

推荐

0

反对

« 上一篇: [图形数据库Neo4J简介](#)
» 下一篇: [多变量微积分笔记20——球坐标系](#)

posted on 2018-05-20 13:44 我是8位的 阅读(1644) 评论(0) 编辑 收藏 举报

[刷新评论](#) [刷新页面](#) [返回顶部](#)

登录后才能查看或发表评论，立即 [登录](#) 或者 [逛逛](#) 博客园首页

【推荐】华为 HWD 2022 故事征集，分享最打动你的科技女性故事
【推荐】华为开发者专区，与开发者一起构建万物互联的智能世界

广告 X

yo zodcs.com

永中DCS_文档在线
预览解决方案

兼容不同版本的Office、PDF等，特定格式的合作，支持Windows、信创环境下一键部署

打开 >

编辑推荐:

- 革命性创新，动画杀手铜 @scroll-timeline
- 戏说领域驱动设计（十二）—— 服务
- ASP.NET Core 6框架揭秘实例演示[16]: 内存缓存与分布式缓存的使用
- .Net Core 中无处不在的 Async/Await 是如何提升性能的?
- 分布式系统改造方案 —— 老旧系统改造篇

#她的梦想在发光#
HWD科技女性故事有奖征集
活动时间: 2022年3月8日~3月18日

最新新闻:

- 乔布斯的创业搭档：他缺乏工程师才能，不得不锻炼营销能力来弥补
 - 美国大厂码农薪资曝光：年薪18万美元，够养家，不够买海景房
 - 两张照片就能转视频！Google提出FLIM帧插值模型
 - Android 再推“杀手级”功能，可回收 60% 存储空间
 - 溺在理财暴雷潮的投资人：本金63万，月兑25元不够卖菜
- » 更多新闻...

Powered by:
博客园

Copyright © 2022 我是8位的
Powered by .NET 6 on Kubernetes