

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

概率统计14——几何分布

我家小朋友年方1岁半，家里每天上午都要出去遛小孩。现在小朋友有两项爱好，在家翻垃圾桶，出门捡烟头。

翻垃圾桶可以有效地限制，捡烟头可是防不胜防。

也许烟头能散发出特殊的能量波动，小区的绿化带和草坪上的大部分烟头都能被小朋友准确地发现，如果他在不规则的前进路线中突然停下了，那肯定是看到了新的烟头。



错失奖励概率

在我的严密监视下，小朋友捡烟头的几率已经从原来的“绝不放过”下降到了现在的20%，如果他直到回家还没有烟头，就能得到一个棒棒糖作为奖励。当然，每个烟头对他来说仍然存在着独特的魅力，在最初遇到的3个烟头中没有经受住考验的概率是多少？

设随机变量X是小朋友最终捡起烟头时所发现的烟头个数，那么：

$$P(X = 1) = P(\text{捡起了遇到的第 1 个烟头}) = 0.2$$

$$\begin{aligned} P(X = 2) &= P(\text{捡起了遇到的第 2 个烟头}) \\ &= P(\text{放过了第 1 个烟头})P(\text{捡起第 2 个烟头}) \\ &= 0.8 \times 0.2 \\ &= 0.16 \end{aligned}$$

$$P(X = 3) = 0.8 \times 0.8 \times 0.2 = 0.128$$

最初遇到的3个烟头中没有抵抗住诱惑的概率可以用 $P(X \leq 3)$ 表示：

$$P(X \leq 3) = P(X = 1) + P(X = 2) + P(X = 3) = 0.448$$

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°
--猫猫猫猫猫大人

每个烟头都是一次考验，我的关注点是，小朋友在第几次考验时会捡起烟头？

质量函数

小朋友每次遇到烟头的行为都是一个独立的随机试验，试验只有成功和失败两种结果，且对于每个随机试验来说，成功的概率都是相同的。对于小朋友自己来说，捡起烟头是成功，错过才是失败，只要成功一次就没有棒棒糖，试验结束。现在用p表示成功率（捡起烟头的概率）， $q = 1 - p$ 表示失败率，随机变量X表示第一次成功时所经历的试验次数，那么在进行了r次试验后才遇到第一次成功的概率（或者说在第r次试验取得成功前，需要经历r-1次失败的概率）可以表示为：

$$P(X = r) = pq^{r-1}, \quad r = 1, 2, 3, \dots$$

这个分布就是几何分布（Geometric distribution），X服从几何分布，记为 $X \sim GE(p)$ 。

下面的python代码展示了r取不同值时的 $P(X=r)$ 。

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3 from scipy import stats
4
5 p = 0.2 # 成功率
6 print('X ~ GE({0})'.format(p))
7
8 rs = np.array(range(1, 11, 1)) # 随机变量的取值
9 # 几何分布 X ~ GE(0.2)
10 ps = stats.geom.pmf(rs, p) # 每个随机变量对应的概率
11 for i, r in enumerate(rs):
12     print('P(X={0})={1}'.format(r, ps[i]))
13
14 plt.bar(left=rs, height=ps, width=0.5)
15 plt.xlabel('r, X=r表示第r次试验才成功')
16 plt.ylabel('P(X=r)')
17 plt.rcParams['font.sans-serif'] = ['SimHei'] # 用来正常显示中文标签
18 plt.show()
```

- X ~ GE(0.2)
- P(X=1)=0.2
- P(X=2)=0.16000000000000003
- P(X=3)=0.12800000000000003
- P(X=4)=0.10240000000000003
- P(X=5)=0.08192000000000002
- P(X=6)=0.06553600000000002
- P(X=7)=0.052428800000000025
- P(X=8)=0.041943040000000015
- P(X=9)=0.033554432000000016

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

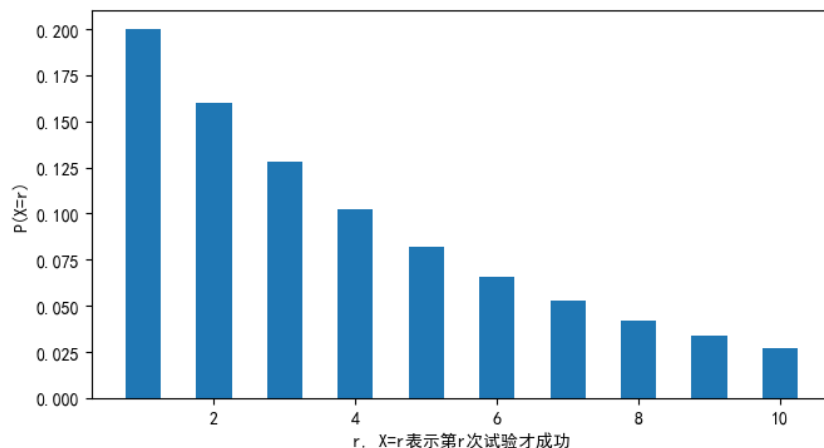
5. Re:寻找“最好” (2) ——欧拉-拉格朗日方程

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

$$P(X=10)=0.026843545600000015$$



对于几何分布来说， r 是大于等于1的自然数，并且只有成功的概率处于 $(0, 1)$ 之间才有意义，因此当 $r = 1$ 时， $P(X = r)$ 达到最大值，随着 r 的增大， $P(X = r)$ 也越来越小。

几何分布是一个典型的长尾分布，第一次就成功的概率是最高的，这看似有违直觉。这里需要注意的是，“第 r 次才取得成功”和“在 r 次之内取得成功”是两回事，后者才是概率越来越大。

实际中有不少随机变量服从几何分布，例如某产品的不合格率为0.05，则首次查到不合格品的检查次数 $X \sim \text{GE}(0.05)$ 。

分布的形状

现在回到问题的关注点：小朋友在第几次考试时会捡起烟头？这个问题并不能确切地回答，可以回答的是，在 r 次考试之内捡起烟头的概率。这个概率可以用分布函数表示：

$$F(r) = P(X \leq r) = \sum_{i=1}^r P(X = i)$$

另一种计算方法是：

$$F(r) = P(X \leq r) = 1 - P(X > r) = 1 - q^r, \quad r = 1, 2, 3, \dots$$

$P(X > r)$ 表示在取得第一次成功时，前 r 次试验都失败的概率。 $F(r)$ 更为专业的说法是：为得到1次成功而进行 r 次伯努利试验， r 的概率分布。

下面的代码展示了几何分布的形状：

```
1 fs = stats.geom.cdf(rs, p) # 每个r对应的分布
2 for i, r in enumerate(rs):
3     print('F({0})=P(X<={0})={1}'.format(r, fs[i]))
4
5 plt.bar(left=rs, height=fs, width=0.5)
6 plt.xlabel('r, X<=r表示前r次试验至少有一次成功')
7 plt.ylabel('F(r) = P(X <= r)')
8 plt.rcParams['font.sans-serif'] = ['SimHei'] # 用来正常显示中文标签
9 plt.show()
```

$$F(1)=P(X \leq 1)=0.19999999999999998)$$

$$F(2)=P(X \leq 2)=0.36)$$

$$F(3)=P(X \leq 3)=0.488)$$

$$F(4)=P(X \leq 4)=0.5904)$$

$$F(5)=P(X \leq 5)=0.67232)$$

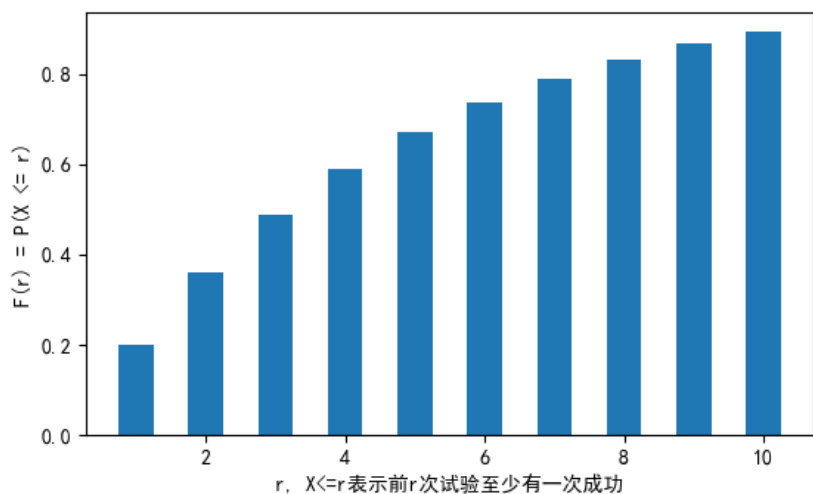
$$F(6)=P(X \leq 6)=0.7378560000000001)$$

$$F(7)=P(X \leq 7)=0.7902848)$$

$$F(8)=P(X \leq 8)=0.83222784)$$

$$F(9)=P(X \leq 9)=0.865782272)$$

$$F(10)=P(X \leq 10)=0.8926258176)$$



小朋友大约有89%的概率在10次考验内捡起烟头，他也因此很少得到棒棒糖。

期望和方差

$X \sim \text{GE}(p)$, $q=1-p$, $P(X=r) = pq^{r-1}$, 当 $r \rightarrow \infty$ 时:

$$E[X] = \frac{1}{p}, \quad \text{Var}(X) = \frac{1-p}{p^2} = \frac{q}{p^2}$$

来看看是怎么得出的。

期望

$$\begin{aligned} E[X] &= \sum_{r=1}^{\infty} rP(X=r) = \sum_{r=1}^{\infty} r p q^{r-1} = p \sum_{r=1}^{\infty} r q^{r-1} \\ (q^r)' &= r q^{r-1} \\ \Rightarrow E[X] &= p \sum_{r=1}^{\infty} r q^{r-1} = p \left(\sum_{r=1}^{\infty} q^r \right)' \end{aligned}$$

可以根据几何级数的公式继续计算:

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}, \quad (|x| < 1)$$

$$\Rightarrow \sum_{r=1}^{\infty} q^r = \left(\sum_{r=0}^{\infty} q^r \right) - q^0 = \frac{1}{1-q} - 1 = \frac{q}{1-q}$$

$$\left(\sum_{r=1}^{\infty} q^r \right)' = \left(\frac{q}{1-q} \right)' = \frac{1}{(1-q)^2} \quad \text{①}$$

求导的除法法则

$$E[X] = p \left(\sum_{r=1}^{\infty} q^r \right)' = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}$$

对于几何分布来说, $p=0.2$ 表示单次试验成功的概率是0.2,
 $E[X]=1/p=5$ 是在告诉我们, 期望在第5次试验时获得成功, 或者说5次试验中就有一次趋向成功。

方差

根据 $\text{Var}(X) = E[X^2] - E[X]^2$ 来计算几何分布的方差:

$$E[X^2] = \sum_{r=1}^{\infty} r^2 P(X=r) = \sum_{r=1}^{\infty} r^2 p q^{r-1} = p \sum_{r=1}^{\infty} r^2 q^{r-1} = p \left(\sum_{r=1}^{\infty} r q^r \right)'$$

$$= p \left(q \sum_{r=1}^{\infty} r q^{r-1} \right)' = p \left(q \left(\sum_{r=1}^{\infty} q^r \right)' \right)'$$

将①代入:

$$E[X^2] = p \left(q \left(\sum_{r=1}^{\infty} q^r \right)' \right)' = p \left(\frac{q}{(1-q)^2} \right)' = p \frac{1+q}{(1-q)^3} = p \frac{1+(1-p)}{p^3} = \frac{2-p}{p^2}$$

$$\text{Var}(X) = E[X^2] - E[X]^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2} = \frac{q}{p^2}$$

出处: 微信公众号 "我是8位的"

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!

扫描二维码关注作者公众号 "我是8位的"



随笔

分类: 概率

标签: 公式推导, 方差, 几何分布, 期望

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: [双色球的概率和收益](#)

» 下一篇: [概率统计15——泊松分布](#)

posted on 2020-01-17 16:16

我是8位的

阅读(2333)

评论(0)

编辑

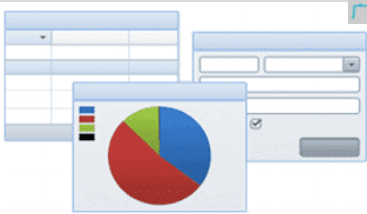
收藏

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



专业WebUI控件库, 快速开发企业管理软件

miniui.com [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...

历史上的今天:
2018-01-17 多变量微积分笔记2——最小二乘法

Powered by:
[博客园](#)
Copyright © 2022 我是8位的
Powered by .NET 6 on Kubernetes