

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

概率统计13——二项分布与多项分布

原文 | <https://mp.weixin.qq.com/s/bOchsmHTINKKlyabCQKMSg>

相关阅读

[最大似然估计\(概率10\)](#)
[寻找“最好” \(3\) 函数和泛函的拉格朗日乘数法](#)

伯努利分布

如果随机试验仅有两个可能的结果，那么这两个结果可以用0和1表示，此时随机变量X将是一个0/1的变量，其分布是单个二值随机变量的分布，称为伯努利分布。注意伯努利分布关注的是结果只有0和1，而不管观测条件是什么。

性质

设p是随机变量等于1的概率，伯努利分布有一些特殊的性质：

$$P(X = 1) = p, \quad P(X = 0) = 1 - p$$

将上面的两个式子合并：

$$P(X = i) = p^i(1 - p)^{1-i}, \quad i = 0, 1$$
$$\text{if } i = 0, \quad \text{then } P(X = 0) = p^0(1 - p)^{1-0} = 1 - p$$
$$\text{if } i = 1, \quad \text{then } P(X = 0) = p^1(1 - p)^{1-1} = p$$

伯努利变量是离散型，并且是一个0/1变量，它的数学期望是：

$$E[X] = \sum_i x_i P(x_i) = x_0 P(x_0) + x_1 P(x_1) = 0 + P(1) = p$$
$$\text{其中 } x_0 = 0, \quad x_1 = 1$$

方差是：

$$\text{Var}(X) = p(1 - p)$$

极大似然

最大似然估计(概率10)

对于伯努利分布的质量函数来说，p是唯一的参数。如果给定N个独立同分布的样本 $\{x^{(1)}, x^{(2)}, \dots, x^{(N)}\}$ ， $x^{(t)}$ 是投硬币的结果，是随机变量， $x^{(t)} \in \{0, 1\}$ ，可以通过极大似然估计，根据样本推测出p的取值：

$$L(p) = \prod_{i=1}^N P(X = x^{(i)}) = \prod_{i=1}^N p^{x^{(i)}}(1 - p)^{1-x^{(i)}}$$

取对数似然函数：

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

$$\begin{aligned}\ln L(p) &= \ln \prod_{i=1}^N p^{x^{(i)}} (1-p)^{1-x^{(i)}} \\ &= \sum_{i=1}^N \ln p^{x^{(i)}} (1-p)^{1-x^{(i)}} \\ &= \sum_{i=1}^N \left(\ln p^{x^{(i)}} + \ln (1-p)^{1-x^{(i)}} \right) \\ &= \sum_{i=1}^N \left(x^{(i)} \ln p + (1-x^{(i)}) \ln (1-p) \right) \\ &= \sum_{i=1}^N x^{(i)} \ln p + \sum_{i=1}^N (1-x^{(i)}) \ln (1-p) \\ &= \sum_{i=1}^N x^{(i)} \ln p + \sum_{i=1}^N \ln (1-p) - \sum_{i=1}^N x^{(i)} \ln (1-p) \\ &= \sum_{i=1}^N x^{(i)} \ln p + N \ln (1-p) - \sum_{i=1}^N x^{(i)} \ln (1-p)\end{aligned}$$

$$\begin{aligned}\frac{\partial \ln L(p)}{\partial p} &= \frac{\partial}{\partial p} \sum_{i=1}^N x^{(i)} \ln p + \frac{\partial}{\partial p} N \ln (1-p) - \frac{\partial}{\partial p} \sum_{i=1}^N x^{(i)} \ln (1-p) \\ &= \sum_{i=1}^N x^{(i)} \frac{1}{p} - \frac{N}{1-p} + \sum_{i=1}^N x^{(i)} \frac{1}{1-p} \\ &= \frac{1}{p} \sum_{i=1}^N x^{(i)} - \frac{N}{1-p} + \frac{1}{1-p} \sum_{i=1}^N x^{(i)} \\ &= 0\end{aligned}$$

$$\Rightarrow \frac{1}{p} \sum_{i=1}^N x^{(i)} = \frac{N}{1-p} - \frac{1}{1-p} \sum_{i=1}^N x^{(i)}$$

等式两侧同时乘以 $\mu(1-p)$ ： $(1-p) \sum_{i=1}^N x^{(i)} = pN - p \sum_{i=1}^N x^{(i)}$

$$\Rightarrow \hat{p} = \frac{1}{N} \sum_{i=1}^N x^{(i)}$$

这是个符合直觉的结果，即使没学过概率和极大似然也能得出这个结论。

二项分布

假设某个试验是伯努利试验，成功概率用p表示，那么失败的概率为1-p。现在进行了N次这样的试验，成功了x次，失败了N-x次，发生这种情况的概率是多少？

质量函数

对于每次实验来说，成功的概率都是p，失败的概率是1-p。假设已经完成了N次试验，并且前x次都成功了，后N-x次都失败了：

$$\begin{aligned}&\underbrace{\text{成功, 成功, \dots, 成功}}_{x \text{ 次}}, \underbrace{\text{失败, 失败, \dots, 失败}}_{1-x \text{ 次}} \\ P(N \text{ 次试验中前 } x \text{ 次成功}) &= P(\text{前 } x \text{ 次成功})P(\text{后 } N-x \text{ 次失败}) = p^x (1-p)^{N-x}\end{aligned}$$

x次成功的情况当然不止一种，比如成功和失败交叉在一起：

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

成功, 失败, 成功, ..., 成功, 失败
共x次成功, 1-x次失败

$$P(\text{成功, 失败, 成功, } \dots, \text{成功, 失败}) = p \times (1 - p) \times p \times \dots \times p \times (1 - p) = p^x(1 - p)^{N-x}$$

这种成功和失败的排列顺序共有 C_N^x 种不同的情况，因此对于任意N次伯努利试验，成功了x次的概率是：

$$P(x) = C_N^x p^x (1 - p)^{N-x}$$

C_N^x 的另一种记法是 $\binom{N}{x}$ 。

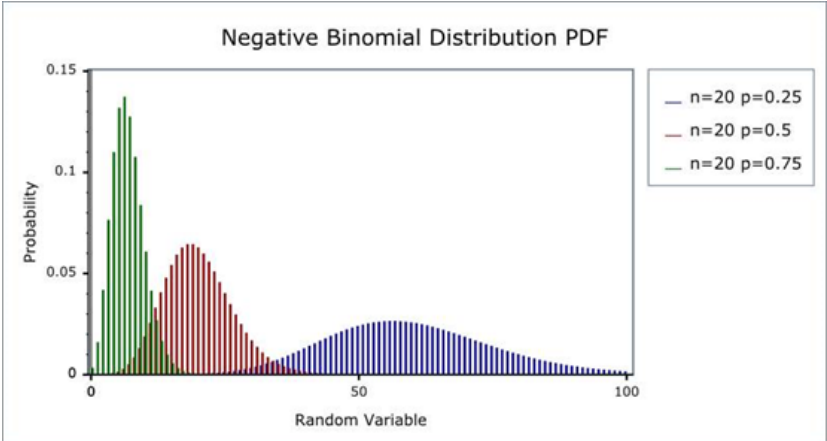
$$P(x) = C_N^x p^x (1 - p)^{N-x} = \binom{N}{x} p^x (1 - p)^{N-x} = \frac{N!}{x!(N-x)!} p^x (1 - p)^{N-x}$$

P(x)就是二项分布的质量函数，是N次伯努利试验中取得x次成功的概率。

性质

二项分布的均值和方差分别为Np和Np(1-p)。

从二项分布的质量函数P(x)可知，概率分布只与试验次数N和成功概率p有关，p越接近0.5，二项分布将越对称。保持二项分布试验的次数N不变，随着成功概率p逐渐接近0.5，二项分布逐渐对称，且近似于均值为Np、方差为Np(1-p)的正态分布：



多项分布

多项分布是二项分布的扩展，其中随机试验的结果不是两种状态，而是K种互斥的离散状态，每种状态出现的概率为 p_i ， $p_1 + p_1 + \dots + p_K = 1$ ，在这个前提下共进行了N次试验，用 $x_1 \sim x_K$ 表示每种状态出现次数， $x_1 + x_2 + \dots + x_K = N$ ，称 $X=(x_1, x_2, \dots, x_K)$ 服从多项分布，记作 $X \sim PN(N: p_1, p_2, \dots, p_n)$ 。

质量函数

如果说二项分布的典型案例是扔硬币，那么多项分布就是扔骰子。骰子有6个不同的点数，扔一次骰子，每个点数出现的概率（对应 $p_1 \sim p_6$ ）都是1/6。重复扔N次，6点出现x次的概率是：

$$P(x) = C_N^x p_6^x (1 - p_6)^{N-x} = \binom{N}{x} p_6^x (1 - p_6)^{N-x}$$

这和二项分布的质量函数类似。现在将问题扩展一下，扔N次骰子，1~6出现次数分别是 $x_1 \sim x_6$ 时的概率是多少？

仍然和二项式类似，假设前 x_1 次都是1点，之后的 x_2 次都是2点.....最后 x_6 次都是6点：

$$\underbrace{1, 1, \dots, 1}_{x_1 \text{次}} \underbrace{2, 2, \dots, 2}_{x_2 \text{次}} \dots \underbrace{6, 6, \dots, 6}_{x_6 \text{次}}$$

$$\begin{aligned} P(\text{前 } x_1 \text{ 次都是 1 点, 之后的 } x_2 \text{ 次都是 2 点 } \dots \text{ 最后 } x_6 \text{ 次都是 6 点}) \\ = P(\text{前 } x_1 \text{ 次都是 1 点, }) P(\text{之后的 } x_2 \text{ 次都是 2 点}) \dots P(\text{最后 } x_6 \text{ 次都是 6 点}) \\ = p_1^{x_1} p_2^{x_2} \dots p_6^{x_6} \end{aligned}$$

1~6出现次数分别是 $x_1 \sim x_6$ 的情况不止一种，1点出现 x_1 次的情况有 $C_N^{x_1}$ 种；在1点出现 x_1 次的前提下，2点出现 x_2 次的情况有 $C_{N-x_1}^{x_2}$ 种；在1点出现 x_1 次且2点出现 x_2 次的前提下，3点出现 x_3 的情况有 $C_{N-x_1-x_2}^{x_3}$ 种.....扔N次骰子，1~6出现次数分别是 $x_1 \sim x_6$ 时的概率是：

$$\begin{aligned} P(x_1, x_2, \dots, x_6) &= C_N^{x_1} p_1^{x_1} C_{N-x_1}^{x_2} p_2^{x_2} \dots C_{N-x_1-x_2-\dots-x_5}^{x_6} p_6^{x_6} \\ &= C_N^{x_1} C_{N-x_1}^{x_2} \dots C_{N-x_1-x_2-\dots-x_5}^{x_6} p_1^{x_1} p_2^{x_2} \dots p_6^{x_6} \end{aligned}$$

根据①：

$$\begin{aligned} C_N^x &= \frac{N!}{x!(N-x)!} \\ C_N^{x_1} C_{N-x_1}^{x_2} \dots C_{N-x_1-x_2-\dots-x_5}^{x_6} \\ &= \frac{N!}{x_1!(N-x_1)!} \times \frac{(N-x_1)!}{x_2!(N-x_1-x_2)!} \times \dots \times \frac{(N-x_1-\dots-x_5)!}{x_6!(N-x_1-\dots-x_6)!} \\ &\quad \text{每一项的分子都能和后一项的分母约分} \\ &= \frac{N!}{x_1! x_2! \dots x_6! (N-x_1-\dots-x_6)!} \\ &= \frac{N!}{x_1! x_2! \dots x_6! 0!} \quad (0! = 1) \\ &= \frac{N!}{x_1! x_2! \dots x_6!} \end{aligned}$$

最终，扔骰子的概率质量函数是：

$$P(x_1, x_2, \dots, x_6) = \frac{N!}{x_1! x_2! \dots x_6!} p_1^{x_1} p_2^{x_2} \dots p_6^{x_6}$$

把这个结论推广到多项分布：某随机实验如果有K种可能的结果 $C_1 \sim C_K$ ，它们出现的概率是 $p_1 \sim p_K$ 。在N随机试验的结果中，分别将 $C_1 \sim C_K$ 的出现次数记为随机变量 $X_1 \sim X_K$ ，那么 C_1 出现 x_1 次、 C_2 出现 x_2 次..... C_K 出现 x_K 次这种事件发生的概率是：

$$P(X_1 = x_1, X_2 = x_2, \dots, X_K = x_K) = \frac{N!}{x_1! x_2! \dots x_K!} p_1^{x_1} p_2^{x_2} \dots p_K^{x_K} = N! \prod_{i=1}^K \frac{p_i^{x_i}}{x_i!}$$

其中 $x_1 + x_2 + \dots + x_K = N$, $p_1 + p_2 + \dots + p_K = 1$ 。

极大似然

多项式的极大似然是指在随机变量 $X_1=x_1, X_2=x_2, \dots, X_K=x_K$ 时，最可能的 $p_1 \sim p_K$ 。

$$L(p_1, p_2, \dots, p_K) = \max_{p_1, p_2, \dots, p_K} L(x_1, x_2, \dots, x_K; p_1, p_2, \dots, p_K)$$

$$\text{s.t. } \sum_{i=1}^K p_i = 1$$

对数极大似然:

$$\begin{aligned} \ln L(p_1, p_2, \dots, p_K) &= \ln N! \prod_{i=1}^K \frac{p_i^{x_i}}{x_i} \\ &= \ln N! + \ln \prod_{i=1}^K \frac{1}{x_i} + \ln \prod_{i=1}^K p_i^{x_i} \\ &= \ln N! + \sum_{i=1}^K \ln \frac{1}{x_i} + \sum_{i=1}^K \ln p_i^{x_i} \\ &= \ln N! + \left(\sum_{i=1}^K \ln 1 - \sum_{i=1}^K \ln x_i \right) + \sum_{i=1}^K \ln p_i^{x_i} \\ &= \ln N! - \sum_{i=1}^K \ln x_i + \sum_{i=1}^K \ln p_i^{x_i} \end{aligned}$$

现在问题变成了求约束条件下的极值:

$$\begin{aligned} \max_{p_1, p_2, \dots, p_K} \ln L(p_1, p_2, \dots, p_K) \\ \text{s.t. } g(p_1, p_2, \dots, p_K) = \sum_{i=1}^K p_i = 1 \end{aligned}$$

根据拉格朗日乘子法:

寻找“最好” (3) 函数和泛函的拉格朗日乘数法

$$\begin{aligned} \nabla \ln L &= \left\langle \frac{\partial \ln L}{\partial p_1}, \frac{\partial \ln L}{\partial p_2}, \dots, \frac{\partial \ln L}{\partial p_K} \right\rangle, \quad \nabla g = \left\langle \frac{\partial g}{\partial p_1}, \frac{\partial g}{\partial p_2}, \dots, \frac{\partial g}{\partial p_K} \right\rangle \\ \nabla \ln L &= \lambda \nabla g \\ \begin{cases} \frac{\partial \ln L}{\partial p_i} = \frac{\partial}{\partial p_i} \left(\ln N! - \sum_{i=1}^K \ln x_i + \sum_{i=1}^K \ln p_i^{x_i} \right) = \frac{\partial}{\partial p_i} \sum_{i=1}^K \ln p_i^{x_i} = \frac{\partial}{\partial p_i} \ln p_i^{x_i} = \frac{x_i p_i^{x_i-1}}{p_i^{x_i}} = \frac{x_i}{p_i} \\ \frac{\partial g}{\partial p_i} = \frac{\partial}{\partial p_i} \sum_{i=1}^K p_i = 1 \end{cases} \\ \frac{\partial \ln L}{\partial p_i} &= \lambda \frac{\partial g}{\partial p_i} \\ \frac{x_i}{p_i} &= \lambda \\ p_i &= \frac{x_i}{\lambda} \end{aligned}$$

根据约束条件:

$$\begin{aligned} \sum_{i=1}^K p_i &= \sum_{i=1}^K \frac{x_i}{\lambda} = \frac{1}{\lambda} \sum_{i=1}^K x_i = 1 \\ \Rightarrow \lambda &= \sum_{i=1}^K x_i = N \\ \hat{p}_i &= \frac{x_i}{\lambda} = \frac{x_i}{N} \end{aligned}$$

这也是个符合直觉的结论。面对有N个样本的K分类数据集，当 $p_i = x_i/N$ 时， C_i 类最可能出现 x_i 次。为了这个结论我们却大费周章，也许又有人因此而嘲笑概率简单了.....

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: 二项分布, 多项分布, 极大似然, 公式推导

[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: [贝叶斯决策理论 \(1\)](#)
» 下一篇: [伽玛函数](#)

posted on 2019-12-28 19:20 我是8位的 阅读(2694) 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

cloud.e-iceblue.cn

JAVA Office
文档在线编辑
APIs

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



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