

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

概率笔记8——方差、均方差和协方差

除了数学期望外，方差、均方差、协方差也是重要的数字特征。

方差

方差的代数意义很简单，两个数的方差就是两个数差值的平方，作为衡量实际问题的数字特征，方差有代表了问题的波动性。

方差的意义

甲、乙二人是射击队最优秀的两名选手，教练组用每一枪的得分作为成绩，根据历史数据计算出二人的平均成绩，也就是数学期望，结果是二人的实力相当，平均成绩都是9.5。比赛的日子快要到了，但是本次比赛只有一个名额，派谁出战呢？

一个简单的方法是让二人比赛一次，谁赢了选谁。但是二者实力相当，一场比赛无法评判孰强孰弱，这样做充满了随机性，丧失了考量的综合性，因此教练组不考虑这种方法。

本次比赛夺金的政治意义重大，需要慎重考虑，稳定性成了教练组判断的另一指标。成绩稳定性是通过另一个数字特征判断的，这个特数字特征就是方差。

下表示甲、乙二人在一次比赛中的成绩：

射击轮数	1	2	3	4	5
甲的得分	9.9	10	8.5	9.8	9.2
乙的得分	9.4	9.4	9.5	9.6	9.3

甲是发挥型选手，成绩波动较大，可以打出“超级环”，也会打出大失水准的“低级环”；相反，乙的发挥比较稳定，总是与平均成绩接近。

二者的平均成绩是已经提前计算出来的9.5，这个成绩已经与随机性无关，我们在这里也不去追究9.5是怎么算出来的，只需要借助这个数字特征给出一个供教练组参考指标，该指标描述了谁的波动大，谁更稳定。

每一次射击的成绩均会产生波动，用每一次射击的得分减去平均成绩表示本次波动，得到了下面的数据：

射击轮数	1	2	3	4	5
甲的波动	0.4	0.5	-1	0.3	-0.3
乙的波动	-0.1	-0.1	0	0.1	-0.2

把每次波动累加起来就是整体的波动。由于有正有负，如果直接相加的话会正负抵消，最后二人的波动都是-0.1，得出二人同样稳定的结论，这显然是荒谬的。解决这个问题一个方法是每次波动都取绝对值，另一个方法是取波动的平方，平方不用考虑符号的问题，因此比绝对值更简单。现在用 $(X-E(X))^2$ 表示波动，得到了下面的数据：

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。应该是夹角大于90°
--猫猫猫猫大人

射击轮数	1	2	3	4	5
甲的波动	0.16	0.25	1	0.09	0.09
乙的波动	0.01	0.01	0	0.01	0.04

现在可以计算出二人的总体波动了：

$$\sigma(X_{\text{甲}}) = \sum_{i=1}^5 (x_{\text{甲}i} - E(X_{\text{甲}}))^2 = 0.16 + 0.25 + 1 + 0.09 + 0.09 = 1.59$$
$$\sigma(X_{\text{乙}}) = \sum_{i=1}^5 (x_{\text{乙}i} - E(X_{\text{乙}}))^2 = 0.01 + 0.01 + 0 + 0.01 + 0.04 = 0.07$$

可以看出，乙的波动远远小于甲的波动，说明乙的稳定性更高。

至于具体选择谁出赛，从不同的方面考虑会得出不同的答案，我们的在这里仅仅是将方差作为一个指标给教练组参考。数学本身描述了问题的客观规律，但如何利用客观规律评判是人的事情。

方差与数学期望的关系

在计算方差时，数学期望是必须的，用D(X)表示某个随机变量的方差：

$$D(X) = (X - E(X))^2$$

我们注意到E(X)已经是一个具体的数字，是加权平均值，已经事先通过计算去掉了随机性（比如选手的平均成绩，这是一个稳定的概念，不会因为输掉某场比赛就认为这名选手不行了）；X是某个特定的随机变量，也是一个具体的数值，D(X)仅仅是将x=X放置在了一个函数里，最终也可以对得出定值。根据数学期望的特征，常数的数学期望仍是常数，因此：

$$D(X) = E(D(X))$$

因此我们说，某一个随机变量的方差就是该随机变量函数的数学期望。我们可以利用数学期望的性质继续向下计算：

$$\begin{aligned} D(X) &= E(D(X)) \\ &= E((X - E(X))^2) \\ &= E(X^2 - 2XE(X) + E(X)^2) \\ &= E(X^2) - E(2XE(X)) + E(E(X)^2) \\ &= E(X^2) - E(2X)E(E(X)) + E(X)^2 \\ &= E(X^2) - 2E(X)E(X) + E(X)^2 \\ &= E(X^2) - 2E(X)^2 + E(X)^2 \\ &= E(X^2) - E(X)^2 \end{aligned}$$

所以说某一个变量的方差等于平方的期望减去期望的平方：

$$D(X) = E(X^2) - E(X)^2$$
$$E(X^2) = D(X) + E(X)^2$$

虽然D(X)代表某个特定随机变量的方差，E(D(X)) = D(X)，但E(X²) ≠ X²，E(X²)表示的是随机变量函数g(X)=(X²)的数学期望。

我们可以借助数学期望的计算公式计算随机变量的整体方差（参考上一章内容）：

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

$$\sigma(X) = E(D(X)) = \begin{cases} \sum_i D(x_i)p_i = \sum_i (x_i - E(X))^2 p_i & (\text{离散型}) \\ \int_{-\infty}^{\infty} D(x)f(x) dx = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx & (\text{连续型}) \end{cases}$$

均方差 (标准差)

由于方差是数据的平方，与检测值本身相差太大，人们难以直观地衡量，所以常用均方差代替方差判断数据的波动。

所有样本的方差之和除以样本的个数，再把所得值开根号，所得之数就是这组数据的均方差，也叫标准差，常见的离散型公式：

$$\sigma(r) = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - r)^2}$$

其中r就是随机变量的数学期望，也就是加权平均值，N是样本空间中的一部分数据。

协方差

当舞台转向了多维随机变量时，方差就变成了协方差，这里的“协”是指几个变量的协同相关性。

如果(X, Y)是二维随机变量，且D(X)>0, D(Y)>0，则X,Y的协方差的定义是：

$$COV(X, Y) = (X - E(X))(Y - E(Y))$$

和方差类似，E(X)E(Y)是确定的数学期望，对于某一组确定的变量x=X, y=Y来说，X和Y也是定值，因此协方差可进一步转换为：

$$\begin{aligned} COV(X, Y) &= E((X - E(X))(Y - E(Y))) \\ &= E(XY - XE(Y) - YE(X) + E(X)E(Y)) \\ &= E(XY) - E(XE(Y)) - E(YE(X)) + E(E(X)E(Y)) \end{aligned}$$

由于E(X)E(Y)是定值，因此可以根据数学期望的性质进一步计算：

$$\begin{aligned} \text{let } E(Y) &= C_1, \quad E(X) = C_2 \\ \text{then } E(XE(Y)) &= E(XC_1) = C_1E(X) = E(Y)E(X), \\ E(YE(X)) &= E(YC_2) = C_2E(Y) = E(X)E(Y), \\ E(E(X)E(Y)) &= E(C_2C_1) = C_2C_1 = E(X)E(Y) \\ COV(X, Y) &= E(XY) - E(XE(Y)) - E(YE(X)) + E(E(X)E(Y)) \\ &= E(XY) - E(Y)E(X) - E(X)E(Y) + E(X)E(Y) \\ &= E(XY) - E(Y)E(X) \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

从直观上来看，协方差表示的是两个变量总体误差的期望。

如果两个变量的变化趋势一致，也就是说如果两个变量都大于自身的期望值，那么两个变量之间的协方差就是正值；如果两个变量的变化趋势相

反，即其中一个变量大于自身的期望值，另外一个却小于自身的期望值，那么两个变量之间的协方差就是负值。

当两个变量完全一致时，协方差就变成了方差：

$$COV(X, X) = E(X^2) - E(X)^2 = D(X)$$

这相当于同一个变量的协方差等于方差，自己与自己一定同步，无所谓协同。

协方差的性质：

- (1) $COV(X, Y) = COV(Y, X)$
- (2) $COV(aX, bY) = abCOV(X, Y), \quad a, b \text{ is Constant}$
- (3) $COV(X_1 + X_2, Y) = COV(X_1, Y) + COV(X_2, Y)$

协方差矩阵

协方差只能处理二维问题，对于三维以上数据，就需要计算多个协方差，然后用矩阵将其组织起来，这就是协方差矩阵。

以三维随机变量(X,Y,Z)为例，其协方差矩阵用 Σ 表示：

$$\Sigma = \begin{bmatrix} COV(X, X) & COV(X, Y) & COV(X, Z) \\ COV(Y, X) & COV(Y, Y) & COV(Y, Z) \\ COV(Z, X) & COV(Z, Y) & COV(Z, Z) \end{bmatrix}$$

需要注意的是，协方差矩阵计算的是不同维度之间的协方差，而不是不同样本之间的。

简单来说，协方差矩阵就是两两计算各维度之间的协方差，看看每两个维度之间的相关情况。如果各个变量之间相互独立，那么两个不同维度变量的协方差是0，协方差矩阵就是一个对角矩阵，并且对角线上的每个元素都是该维度的方差：

$$\Sigma = \begin{bmatrix} COV(X, X) & 0 & 0 \\ 0 & COV(Y, Y) & 0 \\ 0 & 0 & COV(Z, Z) \end{bmatrix} = \begin{bmatrix} D(X) & 0 & 0 \\ 0 & D(Y) & 0 \\ 0 & 0 & D(Z) \end{bmatrix}$$

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: [方差](#), [均方差](#), [协方差](#)

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: [概率笔记7——数学期望](#)
» 下一篇: [概率笔记9——大数定律](#)

posted on 2019-06-27 15:31 我是8位的 阅读(20251) 评论(1) 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

MYTHERESA
邂逅新季
全新系列现已登陆
[立即选购](#)

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6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