

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19——中心极限定理

大数定律告诉我们，如果要求得一个随机变量的期望，只需要进行多次重复试验，然后取均值就可以了。然而在使用大数定律时仍然需要小心，因为大数定律并没有明确指出到底需要多少次试验才能充分接近我们所期待的极限。无论实验多少次，我们仍然不能否认存在这样的情况：所抛出的骰子全部是同一点数，尽管这种情况发生的概率很小。

用 Y_n 表示一系列独立同分布的随机变量 $X_1, X_2, \dots, X_n$ 之和，既然 $X_1, X_2, \dots, X_n$ 是随机变量，那么它们的和也是随机的。

一个令人惊奇的事实是， Y_n 的标准化形式总是趋近于正态分布。这意味着随机过程呈现出不确定性的表面下，其实是高度组织化的，所有随机变量最终都归于正态分布。这种现象就是中心极限定理的客观背景。

标准化处理

对于 Y_n 来说，我们知道它的期望和方差：

$$E[Y_n] = E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i] = n\mu$$
$$Var(Y_n) = Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i) = n\sigma^2$$

为了简单起见，可以先假设 $\mu=0, \sigma^2 > 0$ ，这样就可以认为 Y_n 期望等于 X_i 的期望。

方差刻画了单个随机变量相对于均值的波动程度，类似地，我们也想要知道随机变量之和的波动，以便了解“骰子全部都是1点”这种小概率事件发生的几率。问题是，当 $n \rightarrow \infty$ 时， Y_n 的方差是发散的：

$$\lim_{n \rightarrow \infty} n\sigma^2 = \infty$$

随着 n 的增大， Y_n 的分布越来越均匀，在这种情况下讨论 $Var(Y_n)$ 没有任何意义：

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

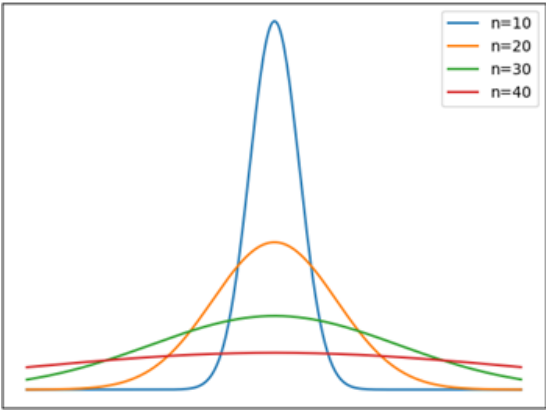
- 1. 隐马尔可夫模型（一）(8)
- 2. 线性代数笔记12——列空间和零空间(7)
- 3. 线性代数笔记3——向量2（点积）(7)
- 4. FP-growth算法发现频繁项集（一）——构建FP树(5)
- 5. 寻找“最好”（2）——欧拉-拉格朗日方程(4)

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



Y_n的分布越来越均匀

为了能够有效地讨论Y_n的方差，需要对其进行标准化处理。

如果把变量做一个线性变换，可以发现下面的现象：

$$E\left[\frac{X-\mu}{\sigma}\right] = \frac{E[X-\mu]}{\sigma} = \frac{E[X]-\mu}{\sigma} = \frac{\mu-\mu}{\sigma} = 0$$
$$Var\left(\frac{X-\mu}{\sigma}\right) = \frac{Var(X-\mu)}{\sigma^2} = \frac{Var(X)}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1$$

只要令X^{*} = (X - μ)/σ，就可以让任意随机变量的期望化为0，方差化为1，这种转换称为标准化或归一化。

Y_n的标准化是：

$$Y_n^* = \frac{Y_n - E[Y_n]}{\sqrt{Var(Y_n)}} = \frac{Y_n - n\mu}{\sqrt{n}\sigma}$$

中心极限定理

中心极限定理告诉我们，如果有一个独立同分布的随机变量序列X₁, X₂, ..., X_n，它们的期望为μ，方差为σ²>0，那么关于这些随机变量之和Y_n的标准化变量Y_n^{*}的分布函数F_n(x)，对于任意x满足：

$$\lim_{n \rightarrow \infty} F_n(x) = \lim_{n \rightarrow \infty} P\{Y_n^* \leq x\} = \lim_{n \rightarrow \infty} P\left\{\frac{Y_n - n\mu}{\sqrt{n}\sigma} \leq x\right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x) \quad (\text{定理一})$$

想要证明这个定理并不容易，但我们可以后文中给出一个从旁侧击的论据。

Y_n^{*}是Y_n的标准化，二者的分布是一回事，因此中心极限定理的另一种写法是：

$$P(a \leq Y_n \leq b) \sim \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

这里的~符号表示“近似于”。也就是说，对于均值为μ，方差为σ²>0的独立同分布的随机变量X₁, X₂, ..., X_n之和Y_n的标准化变量Y_n^{*}，当n充分大时，Y_n^{*}趋近于均值为0，方差为1的正态分布：

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

$$Y_n^* \sim N(0, 1)$$

我们可以把 $\text{Var}(X_i)$ 看作一个微小的误差， Y_n 是微小误差的累加。中心极限定理告诉我们，当无数次微小误差的逐渐累加后，就会形成肉眼可见的巨大变化，并最终接近正态分布。

另外两种表达

在一般的情况下，很难求出 Y_n 的分布函数，尤其是我们面对的经常是“小数”时。在现实生活中，这个“小数”常常就是1，因此才会说“不以成败论英雄”。但是当 n 充分大时，可以通过 $\varphi(x)$ 给出其近似分布，这样就可以利用正态分布对 Y_n 出现的概率进行分析。

把 Y_n^* 的分子和分母同时除以 n ：

$$\frac{(Y_n - n\mu)/n}{\sqrt{n}\sigma/n} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

于是得到了中心极限定理的另一种写法：

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

上一章我们计算了大数定律下均值的期望和方差：

$$E[\bar{X}] = \mu, \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

由此，中心极限定理的第三种写法是：

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

以掷骰子为例，随机变量 X 是每次投掷骰子的结果，出现1~6点的概率都是1/6， $E[X]=3.5$ ，其方差是：

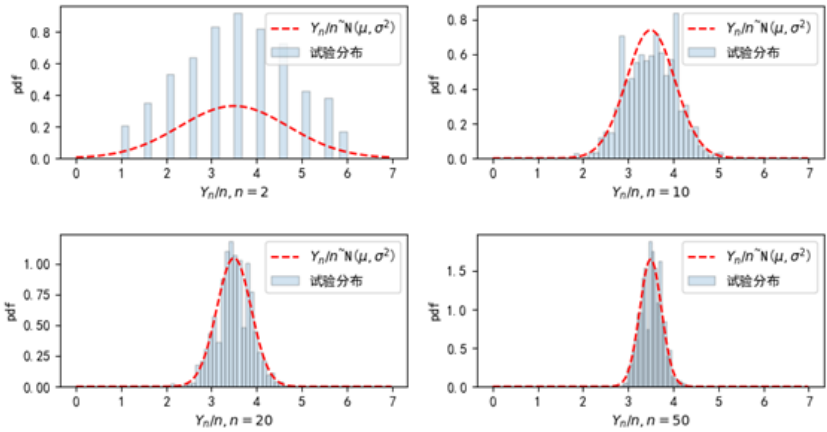
$$\text{Var}(X) = E(X^2) - E(X)^2 = \sum_{k=1}^6 \frac{k^2}{6} - 3.5^2 \approx 2.9167$$

我们用程序模拟多次投掷骰子。每次试验投掷 n 次骰子，试验1000次，看看这种情况下点数均值的分布。



```
1 import numpy as np
2 import matplotlib.pyplot as plt
3 from scipy import stats
4
5 fig = plt.figure(figsize=(10, 5))
6 plt.subplots_adjust(hspace=0.5) # 调整子图之间的上下边距
7
8 mu, sigma_square = 3.5, 2.9167 # 骰子的期望和方差
9 xs = np.arange(0, 7, 0.01)
10 for i, n in enumerate([2, 10, 20, 50]):
11     ax = fig.add_subplot(2, 2, i + 1)
12     means = [np.random.randint(1, 7, n).mean() for i in range(1000)]
13     sigma = np.sqrt(sigma_square / n)
14     ax.hist(means, bins=30, density=True, alpha=0.2, edgecolor='bl
```

```
15  ys = stats.norm.pdf(xs, mu, sigma)
16  ax.plot(xs, ys, 'r--', label='$Y_n/n \sim N(\mu, \sigma^2/n)$')
17  ax.set_xlabel('$Y_n/n, n={}$'.format(n))
18  ax.set_ylabel('pdf')
19  ax.set_title('n={}'.format(n))
20  ax.legend(loc='upper right')
21
22  plt.rcParams['font.sans-serif'] = ['SimHei'] # 用来正常显示中文标签
23  plt.show()
```



n的值越大， Y_n/n 越趋近于 $N(\mu, \sigma^2/n)$ 的正态分布。

示例

浙江大学《概率论与数理统计》中有一个很好的示例，直接贴图一下：

其中 $n=90\ 000, p=1/3$, 即有

$$P\{29\ 500 \leq X \leq 30\ 500\} \approx \Phi\left(\frac{5\sqrt{2}}{2}\right) - \Phi\left(-\frac{5\sqrt{2}}{2}\right) = 0.999\ 5. \quad \square$$

例3 对于一个学生而言,来参加家长会的家长人数是一个随机变量.设一个学生无家长、1名家长、2名家长来参加会议的概率分别为0.05、0.8、0.15.若学校共有400名学生,设各学生参加会议的家长人数相互独立,且服从同一分布.

(1) 求参加会议的家长人数 X 超过450的概率;

(2) 求有1名家长来参加会议的学生人数不多于340的概率.

解 (1) 以 X_k ($k=1, 2, \dots, 400$) 记第 k 个学生来参加会议的家长人数, 则 X_k 的分布律为

X_k	0	1	2
p_k	0.05	0.8	0.15

易知 $E(X_k) = 1.1, D(X_k) = 0.19, k=1, 2, \dots, 400$. 而 $X = \sum_{k=1}^{400} X_k$. 由定理1 随机变量

$$\frac{\sum_{k=1}^{400} X_k - 400 \times 1.1}{\sqrt{400} \sqrt{0.19}} = \frac{X - 400 \times 1.1}{\sqrt{400} \sqrt{0.19}}$$

近似服从正态分布 $N(0, 1)$, 于是

$$\begin{aligned} P\{X > 450\} &= P\left\{\frac{X - 400 \times 1.1}{\sqrt{400} \sqrt{0.19}} > \frac{450 - 400 \times 1.1}{\sqrt{400} \sqrt{0.19}}\right\} \\ &= 1 - P\left\{\frac{X - 400 \times 1.1}{\sqrt{400} \sqrt{0.19}} \leq 1.147\right\} \\ &\approx 1 - \Phi(1.147) = 0.125\ 1. \end{aligned}$$

(2) 以 Y 记有一名家长参加会议的学生人数, 则 $Y \sim b(400, 0.8)$, 由定理

$$\begin{aligned} P\{Y \leq 340\} \\ &= P\left\{\frac{Y - 400 \times 0.8}{\sqrt{400 \times 0.8 \times 0.2}} \leq \frac{340 - 400 \times 0.8}{\sqrt{400 \times 0.8 \times 0.2}}\right\} \end{aligned}$$

• 126 •

$$= P\left\{\frac{Y - 400 \times 0.8}{\sqrt{400 \times 0.8 \times 0.2}} \leq 2.5\right\} \approx \Phi(2.5) = 0.993\ 8.$$

小结

人们在长期实践中认识到频率具有稳定性, 即当试验次数不断增大时, 频率稳定在概率的附近. 这一事实显示了可以用一个数来表征事件发生的可能性的. 这使人想到概率是客观存在的, 进而由频率的性质的启发和抽象给出了概率的定义, 因而频率的稳定性是概率定义的客观基础. 伯努利大数定理则以严密的数学形式论证了频率的稳定性.

中心极限定理表明, 在相当一般的条件下, 当独立随机变量的个数不断增加时, 其分布趋于正态分布. 这一事实阐明了正态分布的重要性. 也揭示了为什么在实际应用中常遇到正态分布, 也就是揭示了产生正态分布变量的源泉. 另一方面, 它提供了独立

随机变量之和 $\sum_{k=1}^n X_k$ (其中 X_k 的方差存在) 的近似分布, 只要和式中加项的个数充分大, 可以不必考虑和式中的随机变量服从什么分布, 都可以用正态分布来近似, 这在应用上是有效和重要的.

中心极限定理的内容包含极限, 因而称它为极限定理是很自然的. 又由于它在统计上的重要性, 称它为中心极限定理, 这是波利亚(Polya)在1920年取的名字.

■ 重要术语及主题

棣莫弗-拉普拉斯定理

我们通常讲的中心极限定理的全称是独立同分布中心极限定理. 棣莫弗-拉普拉斯 (De Moivre-Laplace) 定理可以看作是独立同分布中心极限定理的窄化形式.

设 Y_n ($n=1,2,\dots$) 服从参数为 n , p ($0 < p < 1$) 的二项分布 $Y_n \sim B(n, p)$, Y_n^* 是 Y_n 的标准化, 则对于任意 x , 有:

$$\lim_{n \rightarrow \infty} P\{Y_n^* \leq x\} = \lim_{n \rightarrow \infty} P\left\{\frac{Y_n - np}{\sqrt{np(1-p)}} \leq x\right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x) \quad (\text{定理二})$$

定理二是说, 二项分布随机变量的标准化形式近似于均值为0, 方差为1的正态分布。

根据二项分布的定义可知, Y_n 是 n 重伯努利事件中成功事件发生的次数。因此可以把 Y_n 分解为 n 个伯努利随机变量之和, 我们知道伯努利随机变量的期望是 $\mu = p$, 方差是 $\sigma^2 = p(1-p)$, 因此:

$$\lim_{n \rightarrow \infty} P\left\{\frac{Y_n - np}{\sqrt{np(1-p)}} \leq x\right\} = \lim_{n \rightarrow \infty} P\left\{\frac{Y_n - n\mu}{\sqrt{n}\sigma} \leq x\right\} = \lim_{n \rightarrow \infty} P\{Y_n^* \leq x\}$$

这和定理一的形式相同, 既然相信定理一成立, 那么定理二也成立。

用特例佐证

我们用成功率 $p=1/2$ 的特例来检验棣莫弗-拉普拉斯定理, 此时 $Y_n \sim B(n, 1/2)$, 根据二项分布的性质, 我们也知道这特例条件下 Y_n 的均值、方差和标准差:

$$\mu = np = \frac{n}{2}, \quad \sigma^2 = np(1-p) = \frac{n}{4}, \quad \sigma = \frac{\sqrt{n}}{2}$$

由此可以得到 Y_n 的标准化形式:

$$Y_n^* = \frac{Y_n - n\mu}{\sqrt{n}\sigma} = \frac{Y_n - \frac{n}{2}}{\frac{\sqrt{n}}{2}}$$

进一步可以用 Y_n^* 表示 Y_n :

$$Y_n = \frac{n}{2} + \frac{\sqrt{n}}{2} Y_n^* \quad (1)$$

$$n - Y_n = \frac{n}{2} - \frac{\sqrt{n}}{2} Y_n^* \quad (2)$$

现在进行简单的代数, 假设当 $Y_n = r$ 时, $Y_n^* = x_r$ 。由于 Y_n^* 是 Y_n 的标准化, 因此二者的概率是一样的, 根据二项分布:

$$P(Y_n^* = x_r) = P(Y_n = r) = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r} = \frac{n!}{r!(n-r)!} \times \frac{1}{2^n} \quad (3)$$

现在需要借助一下斯特林公式。

斯特林公式是一种适用于大数阶乘的近似公式。一般来说, 当 n 很大的时候, n 阶乘的计算量十分大, 所以斯特林公式十分好用, 而且, 即使在 n 很小的时候, 斯特林公式的取值也十分准确。

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$

或更精确的:

$$\lim_{n \rightarrow \infty} \frac{\sqrt{2\pi n} \left(\frac{n}{e}\right)^n}{n!} = 1$$

将斯特林公式应用于③：

$$\begin{aligned} \frac{n!}{r!(n-r)!} \times \frac{1}{2^n} &\approx \frac{\sqrt{2\pi n} \left(\frac{n}{e}\right)^n}{\sqrt{2\pi r} \left(\frac{r}{e}\right)^r \sqrt{2\pi(n-r)} \left(\frac{n-r}{e}\right)^{n-r}} \times \frac{1}{2^n} \\ &= \frac{\sqrt{2\pi} \sqrt{nn^n} e^{-n}}{\sqrt{2\pi} \sqrt{rr^r} e^{-r} \sqrt{2\pi} \sqrt{(n-r)} (n-r)^{n-r} e^{-n+r}} \times \frac{1}{2^n} \\ &= \frac{\sqrt{nn^n} e^{-n}}{\sqrt{2\pi} \sqrt{rr^r} \sqrt{(n-r)} (n-r)^{n-r} e^{-n}} \times \frac{1}{2^n} \\ &= \frac{\sqrt{nn^n}}{\sqrt{2\pi} \sqrt{rr^r} \sqrt{(n-r)} (n-r)^{n-r}} \times \frac{1}{2^n} \\ &= \frac{1}{\sqrt{2\pi}} \times \frac{1}{r^r} \times \frac{1}{(n-r)^{n-r}} \times \frac{\sqrt{n}}{\sqrt{r} \sqrt{n-r}} \times \left(\frac{n}{2}\right)^n \\ &= \frac{1}{\sqrt{2\pi}} \times \left(\frac{1}{r}\right)^r \times \left(\frac{1}{n-r}\right)^{n-r} \times \left(\frac{n}{r(n-r)}\right)^{\frac{1}{2}} \times \left(\frac{n}{2}\right)^r \times \left(\frac{n}{2}\right)^{n-r} \\ &= \frac{1}{\sqrt{2\pi}} \times \left(\frac{n/2}{r}\right)^r \times \left(\frac{n/2}{n-r}\right)^{n-r} \times \left(\frac{n}{r(n-r)}\right)^{\frac{1}{2}} \quad ④ \end{aligned}$$

现在将r替换成 x_r ，根据①和②：

$$\begin{aligned} r &= \frac{n}{2} + \frac{\sqrt{n}}{2} x_r, \quad n-r = \frac{n}{2} - \frac{\sqrt{n}}{2} x_r \quad ⑤ \\ \left(\frac{n/2}{r}\right)^r &= \left(\frac{\frac{n}{2}}{\frac{n}{2} + \frac{\sqrt{n}}{2} x_r}\right)^r = \left(\frac{1}{1 + \frac{1}{\sqrt{n}} x_r}\right)^r = \left(1 + \frac{x_r}{\sqrt{n}}\right)^{-r} \\ \left(\frac{n/2}{n-r}\right)^{n-r} &= \left(\frac{\frac{n}{2}}{\frac{n}{2} - \frac{\sqrt{n}}{2} x_r}\right)^{n-r} = \left(\frac{1}{1 - \frac{1}{\sqrt{n}} x_r}\right)^{n-r} = \left(1 - \frac{x_r}{\sqrt{n}}\right)^{-n+r} \\ P(Y_n = r) &= P(Y_n^* = x_r) \approx \underbrace{\frac{1}{\sqrt{2\pi}} \left(\frac{n}{r(n-r)}\right)^{\frac{1}{2}}}_{\alpha} \underbrace{\left(1 + \frac{x_r}{\sqrt{n}}\right)^{-r} \left(1 - \frac{x_r}{\sqrt{n}}\right)^{-n+r}}_{\beta} \end{aligned}$$

x_r 可以用r和n表达，所以上式实际上是关于n和r的函数，由此我们做一个化简：

$$\begin{aligned} \text{let } \alpha(n, r) &= \left(\frac{n}{r(n-r)}\right)^{\frac{1}{2}}, \quad \beta(n, r) = \left(1 + \frac{x_r}{\sqrt{n}}\right)^{-r} \left(1 - \frac{x_r}{\sqrt{n}}\right)^{-n+r} \\ \text{then } P(Y_n = r) &= P(Y_n^* = x_r) \approx \frac{1}{\sqrt{2\pi}} \alpha(n, r) \beta(n, r) \end{aligned}$$

回顾定理二，可以知道：

$$P(Y_n^* = x_r) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_r^2}{2}} \quad ⑥$$

现在 $1/\sqrt{2\pi}$ 已经有了，我们的目标是让 α 和 β 趋近于合理的极限，最终使得二者的积是 $\exp(-x_r^2/2)dt$ 。向这个目标努力，当n很大时，将⑤代入 $\alpha(r, n)$ ：

$$\frac{n}{r(n-r)} = \frac{n}{\left(\frac{n}{2} + \frac{\sqrt{n}}{2}x_r\right)\left(\frac{n}{2} - \frac{\sqrt{n}}{2}x_r\right)} = \frac{n}{\frac{n^2}{4} - \frac{nx_r}{4}} = \frac{4}{n - x_r}$$

$$\lim_{n \rightarrow \infty} \alpha(n, r) = \lim_{n \rightarrow \infty} \left(\frac{n}{r(n-r)} \right)^{\frac{1}{2}} = \lim_{n \rightarrow \infty} \left(\frac{4}{n - x_r} \right)^{\frac{1}{2}} = \frac{2}{\sqrt{n}}$$

$$\alpha(n, r) \approx \frac{2}{\sqrt{n}}$$

得到了这么个结果，似乎作用十分有限。再次利用⑤将会发现一个奇妙的关系：

$$1 = (r+1) - r = \frac{n}{2} + \frac{\sqrt{n}}{2}x_{r+1} - \frac{n}{2} - \frac{\sqrt{n}}{2}x_r = \frac{\sqrt{n}}{2}(x_{r+1} - x_r)$$

$$\Rightarrow x_{r+1} - x_r = \frac{2}{\sqrt{n}}$$

$$\alpha(n, r) = x_{r+1} - x_r = \Delta x$$

α 终于和微分扯上关系了， n 很大时，它将成为⑥中的 dx 。

现在来看 $\beta(n, r)$ 。为了便于操作，我们取对数以便消除幂：

$$\ln \beta(n, r) = \ln \left(1 + \frac{x_r}{\sqrt{n}} \right)^{-r} \left(1 - \frac{x_r}{\sqrt{n}} \right)^{-n+r} = -r \ln \left(1 + \frac{x_r}{\sqrt{n}} \right) + (r-n) \ln \left(1 - \frac{x_r}{\sqrt{n}} \right)$$

既然 n 是一个充分大的数，那么可以判定 $x_r/\sqrt{n} < 1$ 。通过无穷级数的知识，我们知道当 $|x| < 1$ 时 $\ln(1+x)$ 的展开式（来源可参考[单变量微积分30——幂级数和泰勒级数](#)），并利用这一点展开 $\ln \beta$ ：

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots \quad |x| < 1$$

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} + \dots \quad |x| < 1$$

$$\text{let } t = \frac{x_r}{\sqrt{n}}, \quad |t| < 1 \quad \text{then}$$

$$\begin{aligned} \ln \beta(n, r) &= -r \ln(1+t) + (r-n) \ln(1-t) \\ &= -r \left(t - \frac{t^2}{2} + \frac{t^3}{3} - \frac{t^4}{4} + \dots \right) + r \left(-t - \frac{t^2}{2} - \frac{t^3}{3} - \frac{t^4}{4} + \dots \right) - n \ln(1-t) \\ &= r \left(-2t - \frac{2t^3}{3} - \dots \right) - n \left(-t - \frac{t^2}{2} - \frac{t^3}{3} - \frac{t^4}{4} + \dots \right) \end{aligned}$$

当 n 充分大时， $|t| < 1$ ， t 的幂次越高，对结果的影响越小，因此我们抛弃高于 t^2 的项，并再次借助⑤：

$$\begin{aligned} \ln \beta(n, r) &\approx -2rt + nt + \frac{nt^2}{2} \\ &= \frac{-2rx_r}{\sqrt{n}} + \sqrt{n}x_r + \frac{x_r^2}{2} \\ &= \frac{-2x_r}{\sqrt{n}} \left(\frac{n}{2} + \frac{\sqrt{n}}{2}x_r \right) - \sqrt{n}x_r + \frac{x_r^2}{2} \\ &= \sqrt{n}x_r - x_r^2 - \sqrt{n}x_r + \frac{x_r^2}{2} \\ &= -\frac{x_r^2}{2} \\ &\Rightarrow \beta(n, r) \approx e^{-\frac{x_r^2}{2}} \end{aligned}$$

现在终于达到了目的：

$$P(Y_n^* = x_r) \approx \frac{1}{\sqrt{2\pi}} \alpha(n, r) \beta(n, r) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_r^2}{2}} \Delta x$$
$$P(a \leq Y_n^* \leq b) = \sum_{a \leq x \leq b} P(Y_n^* = x) \approx \sum_{a \leq x \leq b} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \Delta x \approx \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$$

虽然这只是一个二项分布的特殊情况，但仍然呈现出清晰的道理：任何大量独立同分布的随机过程都将以一个正态分布为模型，这也是很多分布的曲线都是倒钟形的原因。

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: 棣莫弗-拉普拉斯定理, 中心极限定理, 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

1

推荐

0

反对

« 上一篇: 概率统计18——再看大数定律

» 下一篇: 概率统计20——估计量的评选标准

posted on 2020-02-16 17:55 我是8位的 阅读(4104) 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

#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