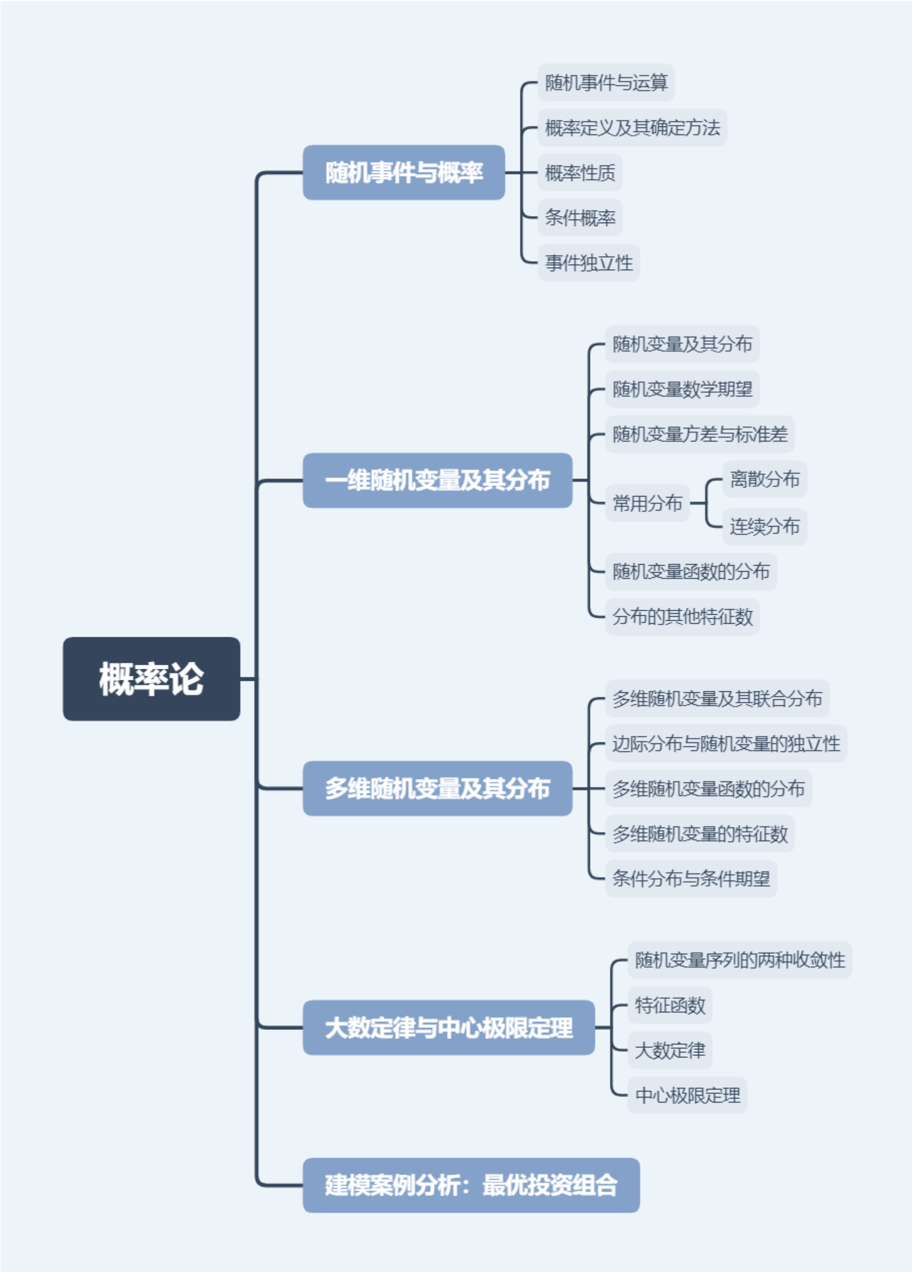
动手学概率论

这是一个关于数学建模相关理论知识的复习专题，同时记录在*Datawhale*社群的第二次打卡活动，从理论到*python*实践，课程参考 *https://github.com/Git-Model/__init_Modeling__*

在复习完高数、线代后，我们再来回顾一下概率论的内容。在大学的时候，我们会有一门《概率论与数理统计》的课程，简单来说，概率论与数理统计研究的对象是随机现象，而概率论主要研究的是随机现象的模型(即概率分布)及其性质，数理统计则是研究随机现象的数据收集、处理及统计推断，本文我们将参考茆诗松的《概率论与数理统计教程》，先来复习概率论部分的知识，主要内容如下：



1. 随机事件与概率

1.1. 随机事件及其运算

1.1.1. 随机现象

在一定条件下,并不总是出现相同结果的现象。如抛一枚硬币与掷一颗骰子，随机现象有两个特点：

- 结果不止一个

- 哪一个结果出现，人们事先不知道。

与之相对的，只有一个结果的现象称为确定性现象，我们前面高数部分的确定性的函数关系就可以用来表示这类现象。对于随机现象，我们如何才能得到某一种可能对应的结果出现的可能性大小呢？最简单的方式就是做实验。对于相同条件下可以重复的随机现象的观察、记录、实验称为随机试验，当然我们也要注意研究不能重复的随机现象。

1.1.2. 样本空间

随机现象的一切可能基本结果组成的集合称为样本空间，记为$\Omega = \{\omega\}$，其中ω表示基本结果，又称为样本点。需要注意的是：

- 样本空间中的元素可以是数也可以不是数(比如硬币的正反)
- 随机现象的样本空间至少有两个样本点(只有一个样本点属于确定性现象)
- 样本点的个数为有限个或者可列的，称为离散样本空间；样本点个数为不可列无限个的，称为连续样本空间(比如掷硬币的结果，一台电视的寿命)

1.1.3. 随机事件与随机变量

随机现象的某些样本点组成的集合称为随机事件，简称事件，通常用$A, B, C, \cdots$表示。如在抛一颗骰子时，“出现奇数点”是一个事件，它是由 1 点、 3 点、 5 点三个基本结果组成，若记这个事件为A， 则有 $A = \{1, 3, 5\}$，它是相应样本空间$\Omega = \{1, 2, 3, 4, 5, 6\}$的一个子集。

由样本空间Ω中的单个元素组成的子集称为基本事件，而样本空间Ω的最大子集(即Ω本身)称为必然事件，样本空间Ω的最小子集(即空集ϕ)表示。

用来表示随机现象结果的变量称为随机变量，常用大写字母X, Y, Z表示，随机变量的含义往往要结合情景。比如：

很多随机现象的结果本身就是数，比如抛一颗骰子，可能出现$1, 2, 3, 4, 5, 6$。若设置$X =$“掷一颗骰子出现的点数”，则$1, 2, 3, 4, 5, 6$就是随机变量X的可能取值，这时

- 事件“出现 3 点”可用“ $X = 3$ ”表示
- 事件“出现点数超过 3 ”可用“ $X > 3$ ”表示
- “ $X \leq 6$ ”是必然事件
- “ $X = 7$ ”是不可能事件

在这个随机现象中，可以再设$Y =$“掷一颗骰子6点出现的次数”，则Y是仅取0或1两个值的随机变量，这时

- “$Y=0$”表示事件“没有出现 6 点”
- “$Y=1$”表示事件“出现 6 点”
- “$Y\le 1$”是必然事件
- “$Y\ge 2$”是不可能事件

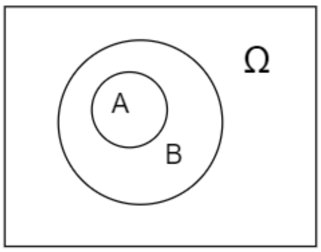
当然还有一些随机现象结果不是数，但是仍然可以设计出有意义的随机变量，比如检验一个产品的可能结果有两个：合格和不合格。我们可以设置$X =$“检查一件产品所得的不合格产品数”，X是仅取0或1两个值的随机变量，且“$X=0$”表示事件“出现合格品”，“$X=1$”表示事件“出现不合格品”。

1.1.4. 事件间的关系

下面的假设在同一个样本空间Ω(即同一个随机现象)中进行，事件间的关系与集合间的关系一样，主要有以下几种：

1. 包含关系

如果属于A的样本点必属于B，则称A被包含于B中，记为$A\subset B$或$B\supset A$，这时事件A发生必然导致事件B发生。对任一事件A，必有$\phi\subset A\subset\Omega$。我们可以用$Venn$图来表示：用一个长方形来表示样本空间$\Omega$，用圆来表示某个事件，当子集$A$中某个样本点出现了，就说事件$A$发生了。

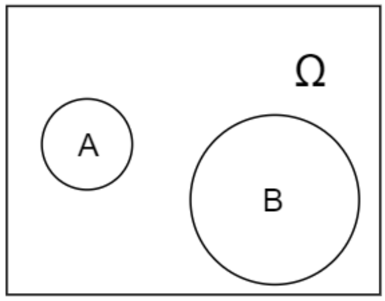


2. 相等关系

如果属于A的样本点属于B，同时属于B的样本点也必属于A，即$A\subset B$且$B\subset A$，那么就称A与B相等，记为$A=B$。

3. 互不相容

如果A与B没有相同的样本点，则称A与B互不相容，从概率论的角度来说，就是事件A与事件B不能同时发生。

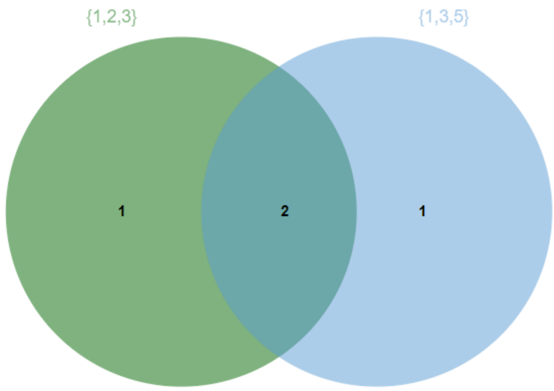


1.1.5. 事件间的运算

事件的运算与集合的运算一致，有并、交、差和余四种运算。

1. 并运算

由事件A与事件B中所有样本点组成的新事件，记为$A \cup B$，此时“A与B中至少有一个会发生”。在掷一颗骰子的试验中，事件A= “出现奇数点” =$\{1,3,5\}$与事件B= “出现点数不超过3“ =$\{1,2,3\}$ 的并为 $A \cup B = 1,2,3,5$。，这里我们可以看到并集是会去重的。



2. 交运算

由事件A与事件B中公共样本点组成的新事件，记为$A \cap B$，此时“A与B同时发生”。对应上面$Venn$图中深绿色的部分。若事件A与事件B互不相容，则其交必为不可能事件，即$AB = \phi$。

3. 差运算

由在事件A中而不在事件B中的样本点组成的新事件，记为$A-B$，此时“事件A发生而事件B不发生”。事件A的对立事件，记为$\bar{A}$，即“由在Ω中而不在A中的样本点组成的新事件”。A与B互为对立的充分必要条件是：$A\cap B=\phi$，且$A\cup B=\Omega$。

1.2. 概率的定义及其确定方法

概率是随机事件发生的可能性大小。虽然随机事件的发生是带有偶然性的，但是其发生的可能性是可以度量的。在概率论发展历史上，曾有过概率的古典定义、概率的几何定义、概率的频率定义来对某一类随机现象定义概率。通过概率的公理化定义，可以给出适合一切随机现象的概率的最一般的定义。

1.2.1. 概率的公理化定义

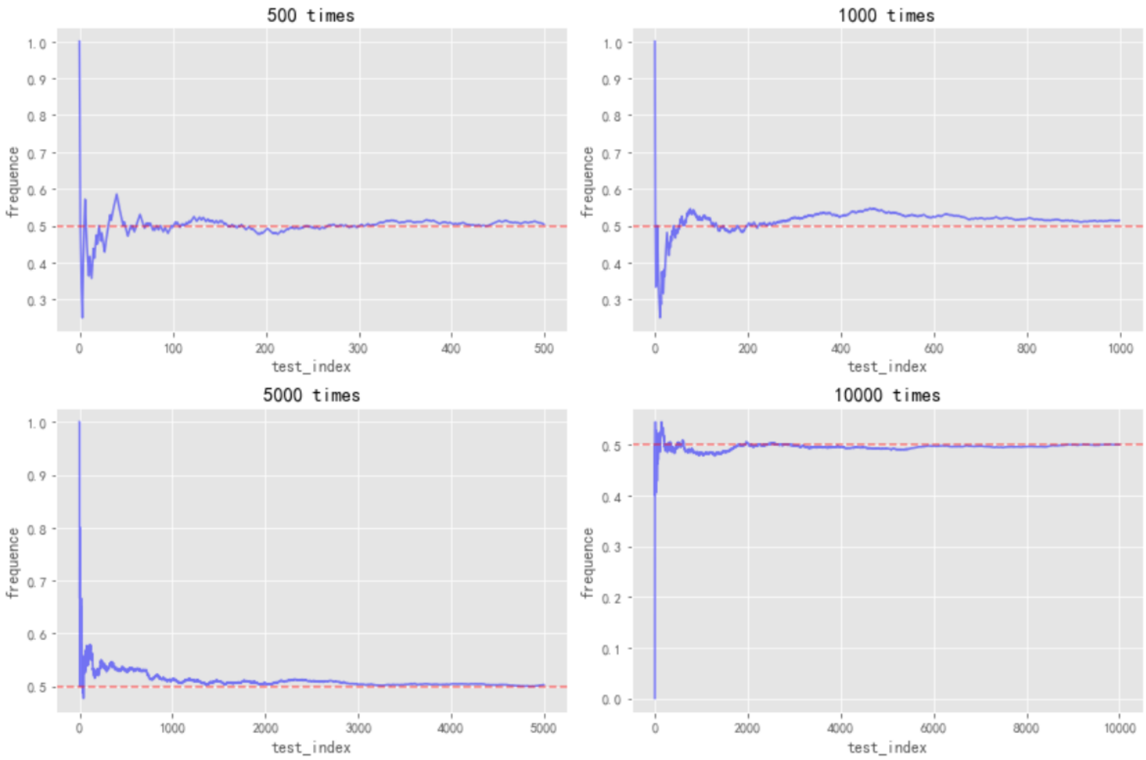
在一个随机现象中，用来表示任意一个随机事件A发生可能性大小的实数，称为概率，记为$P(A)$，其满足：

- 非负性公理： $P(A)\geq 0$
- 正则性公理：样本空间 Ω 的概率 $P(\Omega)=1$
- 可加性公理：若 A_1 与 A_2 互不相容，则有 $P(A_1\cup A_2)=P(A_1)+P(A_2)$

1.2.2. 确定概率的方法

1. 频率法

频率发就是在大量重复实验中，用频率的稳定值当作概率，即$f_n(A)=\dfrac{n(A)}{A}$。我们通过模拟，使用频率来度量n次投掷硬币，正面为1的概率。



2. 古典法

古典法的基本思想如下：

- 所涉及到的随机现象只有有限个样本点，譬如 n 个
- 每个样本点发生的可能性相等(等可能性)
- 若事件 A 含有 k 个样本点，那么事件 A 的概率为

$$P(A)=\frac{\text{事件}A\text{所含样本点的个数}}{\Omega\text{中所有样本点的个数}}=\frac{k}{n}$$

古典概型一般会涉及到排列组合问题，因为我们需要找到样本空间以及事件A中所有的样本数，这里我们通过一个例子来进行说明：假设有n个球，每个球都等可能地被放到N个不同的盒子中的任一个，每个盒子所放球数不同，试求：

（1）指定的$n(n\leq N)$个盒子中各有一球的概率p_1；（2）恰好有$n(n\leq N)$个盒子各有一球的概率。

解析：因为每个球都可放到N个盒子中的任一个且每个盒子可以放多个球，所以n个球的方法共有N^n种。

（1）因为指定了n个盒子放球，所以其实我们只需要考虑n个球在指定的n个盒子中各放一球的放法。第一个球有n种，第二个球有$n-1$种，……，第n个球有1种，这样，总可能数为$n!$，因此，概率$p_1=\dfrac{n!}{N^n}$

（2）因为恰好是有n个盒子中各有一球，所以我们可以分两步：第一步先从N个盒子中取出n个盒子，这样共有C_N^n种取法(组合数)；第二步将n个球放到n个盒子，每个盒子各一球，共有$n!$种取法。因此，概率$p_2 = \frac{C_N^n \cdot n!}{N^n} = \frac{N!}{N^n(N-n)!}$。

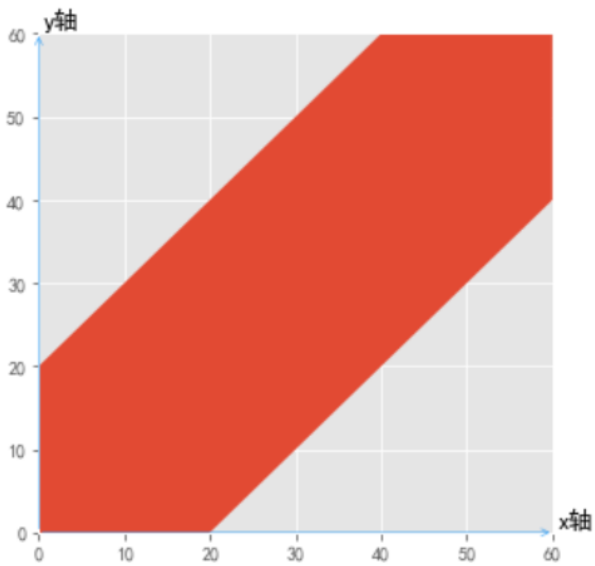
3. 几何法

几何法的基本思想是：

- 如果一个随机现象的样本空间 Ω 能够充满某个区域，其度量(长度、面积或体积等)可用 S_Ω 表示
- 任意一点落在度量相同的子区域内(可能位置不同)是等可能的
- 若事件 A 为 Ω 种的某个子区域，其度量大小可用 S_A 表示，则事件 A 的概率是 $P(A) = \frac{S_A}{S_\Omega}$

我们举一个简单的例子：甲乙两人约定下午6时到7时之间在某处会面，并约定先到者应等候另一个人20min，过时即可离去，试求两人能会面的概率。

分析：以x和y分别表示甲乙两人到达约会地点的时间(以min为单位)，建立直角坐标系。因为两人均是在0至60min内等可能地到达，所以(x, y)的所有可能取值都在正方形区域内，其面积$S_\Omega = 60^2$。而事件A=“亮人能够会面”，相当于两人到达的时间差不超过20分钟，即下图所示的阴影区域



其面积$S_A = 60^2 - 40^2$，所以$P(A) = \frac{S_A}{S_\Omega} = \frac{60^2 - 40^2}{60^2} = \frac{5}{9}$。

1.3. 条件概率

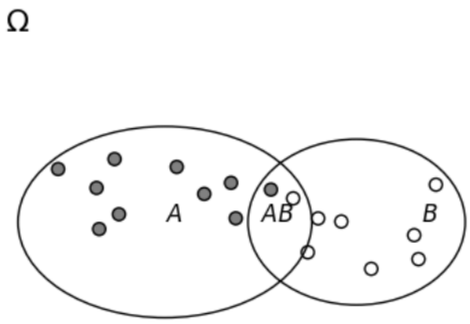
1.3.1. 条件概率公式

条件概率是概率论中重要且实用的概念，是指在某事件B发生的情况下，另一事件A发生的概率，记为$P(A|B)$。前面我们说的事件A的(无条件)概率$P(A)$是两个不同的概念。对于条件概率$P(A|B)$，如果事件B的发生会影响到事件A发生的概率，那么$P(A|B) \neq P(A)$；如果事件B的发生不会影响到事件A发生的概率，也就是我们常说的事件A与事件B相互独立，此时有$P(A|B) = P(A)$。我们还是用一个简单的例子来说明：

有个家庭有两个小孩，其样本空间为$\Omega = bb, bg, gb, gg$。其中b代表男孩，g代表女孩，bg代表大的是男孩，gb代表大的是女孩。那么有

- 事件 A =“家中至少有一个女孩”发生的概率 $P(A) = \frac{3}{4}$
- 已知事件 B =“家中至少有一个男孩”，再求事件 A 发生的概率 $P(A|B) = \frac{2}{3}$

出现不同的原因是，在B发生的条件下，样本空间缩小为$\Omega' = gg, bg, gb$。而在新样本空间中事件A的样本点只有两个，所以$P(A|B) = \frac{2}{3}$。条件概率确切的定义如下：设A与B是样本空间Ω中的两事件，若$P(B) > 0$，则称$P(A|B) = \frac{P(AB)}{P(B)}$为“在$B$发生下$A$的概率”。



1.3.2. 乘法公式

乘法公式是条件概率的第一个重要应用，他从条件概率的角度描述了两个事件A、B同时发生的概率是怎么计算的，具体公式如下：

- 若 $P(B) > 0$，则 $P(AB) = P(B)P(A|B)$

- 若 $P(A_1A_2\cdots A_{n-1})>0$，则
  $P(A_1A_2\cdots A_n)=P(A_1)P(A_2\mid A_1)P(A_3\mid A_1A_2)\cdots P(A_n\mid A_1A_2\cdots A_{n-1})$

例子：一批零件共有 100 个，其中有10个不合格品。从中一个一个取出，求第三次才取得不合格品的概率是多少？

解析：以 A_i 记事件“第 i 次取出的是不合格品”，$i=1,2,3$。则所求概率为 $P\left(\bar{A}_1\bar{A}_2A_3\right)$，由乘法公式得

$$P\left(\bar{A}_1\bar{A}_2A_3\right)=P\left(\bar{A}_1\right)P\left(\bar{A}_2\mid\bar{A}_1\right)P\left(A_3\mid\bar{A}_1\bar{A}_2\right)=\frac{90}{100}\cdot\frac{89}{99}\cdot\frac{10}{98}=0.0826.$$

1.3.3. 全概率公式

全概率公式是计算复杂概率的一个重要方法，使一个复杂概率的计算能化繁为简。设 $B_1,B_2,\cdots,B_n$ 是基本空间 Ω 的一个分割，即$B_1,B_2,\cdots,B_n$互不相容，且 $\overset{n}{\cup}_{i=1}B_i=\Omega$，则对 Ω 中任一事件 A，有$P(A)=\sum\limits_{i=1}^{n}P(A\mid B_i)P(B_i)$。

用一个简单的例子来说明一下，假设在n张彩票中有一张可中奖，求第二人摸到中奖彩票的概率。假设事件A_i表示事件“第i人摸到中奖彩票”，$i=1,2,\cdots,n$。现在要求$P(A_2)$，因为A_1是否发生直接关系到A_2发生的概率，即 $P(A_2|A_1)=0,\quad P(A_2|\bar{A}_1)=\dfrac{1}{n-1}$。而$P(A_1)=\dfrac{1}{n},\quad P(\bar{A}_1)=\dfrac{n-1}{n}$。根据全概率公式：

$$P(A_2)=P(A_1)P(A_2|A_1)+P(\bar{A}_1)P(A_2|\bar{A}_1)=\frac{1}{n}\cdot 0+\frac{n-1}{n}\cdot\frac{1}{n-1}=\frac{1}{n}$$

这说明：摸到中奖彩票的机会与先后次序并无关联。

1.3.4. 贝叶斯公式

在乘法公式和全概率公式的基础上，可以得到贝叶斯公式。假设$B_1,B_2,\cdots,B_n$是样本空间Ω的一个分割，即$B_1,B_2,\cdots,B_n$互不相容，且$\overset{n}{\cup}_{i=1}B_i=\Omega$。如果 $P(A)>0$，$P(B_i)>0$，$i=1,2,\cdots,n$，则：

$$P(B_i|A)=\frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n}P(B_j)P(A|B_j)}$$

这里我们简单推导一下，由条件概率公式$P(B_i|A)=\dfrac{P(AB_i)}{P(A)}$，对分子分母分别用乘法公式、全概率公式，则有

$$P(AB_i) = P(B_i)P(A|B_i)$$
$$P(A) = \sum_{j=1}^{n} P(B_j)P(A|B_j)$$

因此，可以得到贝叶斯公式。

我们来看下面一个例子，某地区居民肝癌发病率为0.0004，现在进行普查(医学研究表明，化验结果是可能存在错误的)。已知患有肝癌的人其化验结果99%呈阳性(有病)，而没患肝癌的人其化验结果99.9%呈阴性(无病)。现某人检查结果呈阳性，试求其真的患病的概率。

解析：设事件B="检查者患有肝癌"，事件A="检查结果呈阳性"，由题可知：

$$\begin{aligned} P(B) &= 0.0004, P(\bar{B}) = 0.9996 \\ P(A|B) &= 0.99, P(A|\bar{B}) = 1 - P(\bar{A}|\bar{B}) = 0.001 \\ P(B|A) &= \frac{P(B)P(A|B)}{P(B)P(A|B) + P(\bar{B})P(A|\bar{B})} \\ &= \frac{0.0004 \times 0.99}{0.0004 \times 0.99 + 0.9996 \times 0.001} = 0.284 \end{aligned}$$

这说明，在检查结果呈阳性的人中，真正患肝癌的不到30%。所以在实际中，常采用复查的方法来减少错误率，譬如对首次检查呈阳性的人群进行复查，此时$P(B) = 0.284$，这时我们再使用贝叶斯公式：

$$P(B|A) = \frac{0.284 \times 0.99}{0.284 \times 0.99 + 0.716 \times 0.001} = 0.997$$

这时检查的正确率大大提高。

需要注意的是：在贝叶斯公式中，如果称$P(B_i)$为B_i的先验概率，那么$P(B_i|A)$就是B_i的后验概率，贝叶斯公式就是专门计算后验概率的。也就是通过A的发生这个新信息，来对B_i的概率做出修正。我们再来举一个例子：狼来了的故事，我们知道前两次村民都是听了小孩子的谎话，但是第三次就不再相信，我们来做一个简单的模拟。

假设事件B="小孩可信"，事件A="小孩说谎"，我们需要求得是$P(B|A)$，也就是对于村民来说小孩说谎可信的概率。不妨设村民对小孩印象很好，$P(B) = 0.8, P(\bar{B}) = 0.2$。

在贝叶斯公式中，我们需要用到$P(A|B)$和$P(A|\bar{B})$，前者表示小孩可信的情况下说谎话的概率，后者表示小孩不可信的情况下说谎话的概率。不妨设$P(A|B) = 0.1, P(A|\bar{B}) = 0.5$，这样第一次小孩说谎的时候，村民知道被骗之后，信任程度发生改变：

$$
\begin{aligned}
P(B|A) &= \frac{P(B)P(A|B)}{P(B)P(A|B)+P(\bar{B})P(A|\bar{B})} \\
&= \frac{0.8\times 0.1}{0.8\times 0.1+0.2\times 0.5} = 0.444
\end{aligned}
$$

在第一次被骗后，村民对小孩的信任度从0.8下降到0.444，在第二次知道被骗后

$$
\begin{aligned}
P(B|A) &= \frac{P(B)P(A|B)}{P(B)P(A|B)+P(\bar{B})P(A|\bar{B})} \\
&= \frac{0.444\times 0.1}{0.444\times 0.1+0.556\times 0.5} = 0.138
\end{aligned}
$$

信任度已经下降都0.138了，如此低的信任度，难怪第三次村民“不上当”了。

2. 随机变量及其分布

在前面我们提到过随机变量，我们把”用来表示随机现象结果的变量“称为随机变量，在这里，我们进一步来探讨关于随机变量的问题。

2.1. 随机变量及其分布

2.1.1. 随机变量的概念

在随机现象中很多样本点本身就是数量表示的(比如骰子点数)，但是也存在样本点本身不是数的情况。此时恶意根据样本点的情况进行赋值(对于一个产品，X表示样本合格情况，$X=0$表示合格品，$X=1$表示不合格品)，则X的取值及其对应概率如下表所示：

X	0	1
P	p	$1-p$

定义在样本空间Ω上的实值函数$X=X(\omega)$称为随机变量，常用大写字母X,Y,Z表示，其取值常用x,y,z表示。

2.1.2. 随机变量的分布函数

随机变量X是样本点ω的一个函数，若B是某些实数组成的集合，即$B \subset \mathbf{R}$，则$\{X \in B\}$表示如下的随机事件：$\{\omega : X(\omega) \in \mathbf{B}\} \subset \Omega$。特别，用等号或者不等号把随机变量$X$与某些实数连接起来，用来表示事件，如$\{X \le a\}$、$\{x > b\}$和$\{a < X < b\}$都是随机事件。

设X是一个随机变量，对任意实数x，称$F(x) = P(X \le x)$为随机变量X的分布函数。任一分布函数$F(x)$都具有如下三条基本性质：

- 单调性 $F(x)$ 是定义在整个实数轴 $(-\infty, +\infty)$ 上的单调非减函数，即对任意的 $x_1 < x_2$，有 $F(x_1) \le F(x_2)$
- 有界性 对任意的 x，有 $0 \le F(x) \le 1$，且 $F(-\infty) = \lim\limits_{x->-\infty} F(x) = 0, F(+\infty) = \lim\limits_{x->+\infty} F(x) = 1$。
- 右连续性 $F(x)$ 是 x 的右连续函数，即对任意的 x_0，有 $\lim\limits_{x->x_0+0} = F(x_0)$。

对任意的实数a与b，有

$$
\begin{aligned}
P(a < X \le b) &= F(b) - F(a) \\
P(X = a) &= F(a) - F(a-0) \\
P(X \ge b) &= 1 - F(b-0) \\
P(X > b) &= 1 - F(b) \\
P(x < b) &= F(b-0) \\
P(a < X < b) &= F(b-0) - F(a) \\
P(a \le X \le b) &= F(b) - F(a-0) \\
P(a \le X < b) &= F(b-0) - F(a-0)
\end{aligned}
$$

2.1.3. 离散随机变量的概率分布列

对于离散随机变量而言，常用以下定义的分布列来表示其分布

设X是一个离散随机变量，如果X的所有可能取值是$x_1, x_2, \cdots, x_n, \cdots,$，则称$X$取$x_i$的概率$p_i = p(x_i) = P(X = x_i)$为$X$的概率分布列，记为$X \sim p_i$。分布列也可以用如下表表示：

$$
\begin{array}{c|ccccc}
X & x_1 & x_2 & \cdots & x_n & \cdots \\
\hline
P & p(x_1) & p(x_2) & \cdots & p(x_n) & \cdots
\end{array}
$$

分布列的基本性质：

- 非负性 $p(x_i) \ge 0,\ i = 1, 2, \cdots$
- 正则性 $\sum\limits_{i=1}^{n} p(x_i) = 1$

由分布列可以写出X的分布函数$F(x)=\sum_{x_i\le x}p(x_i)$

设离散随机变量X的分布列为

X	-1	2	3
P	0.25	0.5	0.25

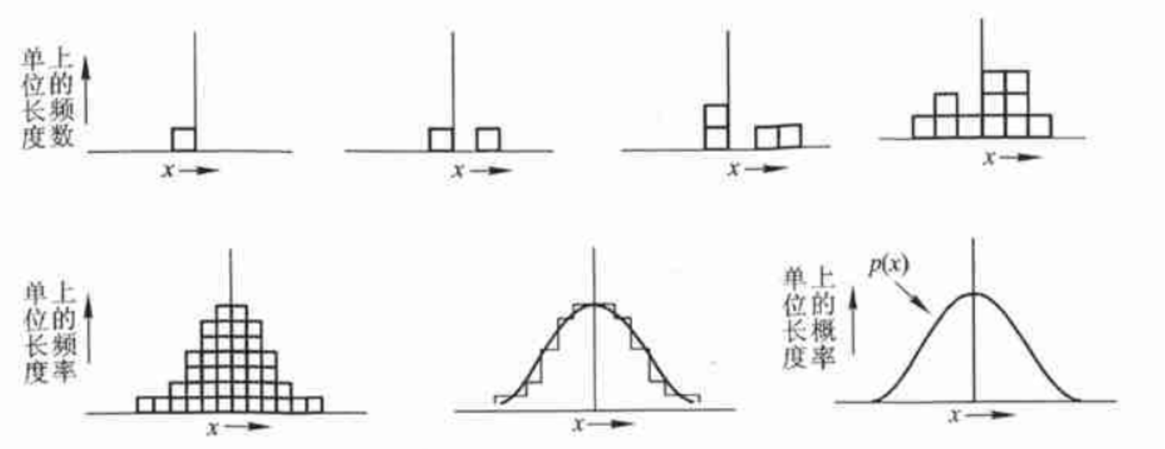
试写出X的分布函数

$$F(x)=\begin{cases}0, & x<-1\\ 0.25, & -1\le x<2\\ 0.25+0.5=0.75, & 2\le x<3\\ 0.25+0.5+0.25=1, & x\ge 3\end{cases}$$

2.1.4. 连续随机变量的概率密度函数

首先我们给个定义，设随机变量X的分布函数为$F(x)$，如果存在实数轴上的一个非负可积函数$p(x)$，使得对任意实数x有$F(x)=\int_{-\infty}^{x}p(t)dt$，则称$p(x)$为$X$的概率密度函数。下面我们再来看看是怎么推导的

加工机械的直径的测量值X是一个连续随机变量。若我们一个接一个地测量轴的直径，把测量值x一个接一个地放到数轴上去，当累积很多测量值x时，就形成一定的图形。为了使这个图形得以稳定，我们把纵轴由“单位长度上的频数”改为“单位长度上的频率”。由于频率的稳定性，随着测量值+的个数越多和单位越小，这个图形就越稳定，其外形就显现出一条光滑曲线(见下图)。这时，这条曲线的纵坐标已是“单位长度上的概率”，当单位长度趋于0时其纵坐标就是“一点上的概率密度”。这时，这条曲线所表示的函数$p(x)$称为概率密度函数，它表示出X“在一些地方(如中部)取值的机会大，在另一些地方(如两侧)取值机会小”的一种统计规律性。



(a) 概率密度函数$p(x)$的形成过程

概率密度函数$p(x)$的值虽不是概率，但乘微分元dx就可得小区间$(x,x+dx)$上概率的近似值，即$p(x)dx \approx P(x<X<x+dx)$。在$(a,b)$上很多相邻的微分元的累加就得到$p(x)$在$(a,b)$上的积分，也就是$X$在$(a,b)$上的概率，即$\int_a^b p(x)dx = P(a<X<b)$，特别地，在$(-\infty,x]$上$p(x)$的积分就是分布函数$F(x)$，即$\int_{-\infty}^{x} p(t)dt = P(X \le x) = F(x)$。同时，在$F(x)$导数存在的点，有$F'(x)=p(x)$，所以我们称$p(x)$是概率密度函数。

2.1.5. 分布列与概率密度函数区别

1. 离散随机变量的分布函数 $F(x)$ 总是右连续的阶梯函数，而连续随机变量的分布函数 $F(x)$ 是整个数轴上的连续函数
2. 离散随机变量 X 在其可能取值的点上的概率不为0，而连续随机变量 X 在 $(-\infty,+\infty)$ 上任一点概率恒为0，这时因为 $P(x=a)=\int_a^a p(x)dx = 0$ 。这表明：不可能事件的概率为0，但概率为0的事件(如 $P(X=a)=0$)不一定是不可能事件。类似地，必然事件的概率为1，但概率为1的事件不一定是必然事件。
3. 由于连续随机变量 X 在仅取一点处的概率恒为0，从而在事件 $\{a \le X \le b\}$ 中剔除端点并不影响其概率，即 $P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = P(a < X < b)$ 。
4. 从(3)可以看出，一个连续分布的密度函数不唯一，如下：

$$p_1(x) = \begin{cases} 1/a, & 0 \le x \le a \\ 0, & \text{其他} \end{cases}, p_2(x) = \begin{cases} 1/a, & 0 \le x \le a \\ 0, & \text{其他} \end{cases}$$

它们都是$(0,a)$上均匀分布的密度函数，但仔细考察这两个函数$p_1(x)$和$p_2(x)$，可以发现

$P(p_1(x) \ne p_2(x)) = P(X=0)+P(X=a)=0$，可见这两个函数在概率意义上是没有差别的，这是概率论与微积分不同之处。

2.2. 随机变量的数学期望

对于一个离散随机变量X，如果其可能取值为$x_1, x2, \cdots, x_n$。若将这n个数相加后除n作为“均值”，则肯定是不妥的。其原因在于X取各个值的概率一般是不同的，概率大的出现的机会就大，则在计算中其权也应该大，因此用取值的概率作为一种“权数”作加权平均是十分合理的。经以上分析,我们就可以给出数学期望的定义：

设离散随机变量X的分布列为$p(x_i)=P(X=x_i)$，$i=1,2,\cdots,n,\cdots$，如果$\sum_{i=1}^{n}|x_i|p(x_i)<\infty$，则称$E(x)=\sum_{i=1}^{n}x_ip(x_i)$为随机变量$X$的数学期望，简称期望或均值。从定义我们可以看到，如果级数$\sum_{i=1}^{n}|x_i|p(x_i)$不收敛，那么$X$的数学期望不存在。

同样的，我们可以得出连续随机变量的期望：设连续随机变量X的概率密度函数为$p(x)$，如果$\int_{-\infty}^{+\infty}|x|f(x)dx<\infty$，则称$E(X)=\int_{-\infty}^{+\infty}xf(x)dx$为$X$的数学期望。同样的，如果$\int_{-\infty}^{+\infty}|x|f(x)dx$不收敛，那么$X$的数学期望不存在。

我们分别举例来说明~

在一个人数为N的人群中普查某种疾病，为此要抽验N个人的血。如果将每个人的血分别检验，则共需检验N次。为了能减少工作量，一位统计学家提出一种方法：按k个人一组进行分组，把同组k个人的血样混合后检验，如果混合血样呈阴性反应，就说明这k个人的血都呈阴性反应，这k个人都无此疾病，因而这k个人只要检验1次就够了，相当于每个人检验$1/k$次，检验的工作量明显减少了。如果混合血样呈阳性反应,就说明这k个人中至少有一人的血呈阳性反应，则再对这k个人的血样分别进行检验，因而这k个人的血要检验$k+1$次，相当于每个人检验$1+1/k$次，这时增加了检验次数。假设该疾病的发病率为p，且得此疾病相互独立。试问此种方法能否减少平均检验次数？

分析：设X为每个人需要检验的次数，则X的分布列为

X	$1/k$	$1+1/k$
P	$(1-p)^k$	$1-(1-p)^k$

则每个人的平均检测次数

$$E(x)=\frac{1}{k}(1-p)^k+\left(1+\frac{1}{k}\right)[1-(1-p)^k]=1-(1-p)^k+\frac{1}{k}。$$

只要使$(1-p)^k>\frac{1}{k}$，就可减少验血次数，而且还可适当选择k使$E(X)$达到最小。譬如，当$p=0.1$时，对不同的k，当$k\geq 34$时，平均验血次数超过1，即比分别检验的工作量还大；而当$k\leq 33$时，平均验血次数在不同程度上得到了减少，特别在$k=4$时，平均验血次数最少，验血工作量可减少 40%。

设X是服从区间(a,b)上的均匀分布，求$E(x)$。均匀分布的概率密度函数为

$$p_1(x)=\begin{cases}\dfrac{1}{b-a}, & a\leq x<b\\ 0, & \text{其他}\end{cases}$$

所以

$$E(X)=\int_a^b \frac{x}{b-a}dx=\frac{1}{b-a}\cdot\frac{x^2}{2}\Big|_a^b=\frac{a+b}{2}$$

实际上，因为X在区间(a,b)上是均匀分布的，所以均值当然就是区间(a,b)的中点，即$(a+b)/2$。

再来简单说说数学期望的性质：

- 若 c 是常数，则 $E(c)=c$ 。
- 对任意常数 a ，有 $E(aX)=aE(X)$ 。
- 设 (X,Y) 是二维随机变量，则有 $E(X+Y)=E(X)+E(Y)$ 。
- 若随机变量X与Y相互独立 , 则有 $E(XY)=E(X)E(Y)$ 。

2.3. 随机变量的方差和标准差

前面我们说到期望可以衡量随机变量X的均值，也就是X总是在$E(X)$附近波动，那如何才能够反映随机变量取值的波动大小呢，比如对于下面两个分布：

$$\begin{array}{c|ccc} X & -1 & 0 & 1 \\ \hline P & 1/3 & 1/3 & 1/3 \end{array} \qquad \begin{array}{c|ccc} Y & -10 & 0 & 10 \\ \hline P & 1/3 & 1/3 & 1/3 \end{array}$$

虽然两个分布的均值都为0，但是很明显Y的取值波动更大，所以我们引入方差和标准差来度量波动大小。

若随机变量X^2的数学期望$E(X^2)$存在，则称偏差平方$(X-E(X))^2$的数学期望$E(X-E(X))^2$为随机变量X的方差，记为

$$Var(X)=E(X-E(X))^2=\begin{cases}\sum_i (x_i-E(X))^2p(x_i), & X\text{为离散随机变量}\\ \int_{-\infty}\infty(x-E(X))^2p(x)dx, & X\text{为连续随机变量}\end{cases}$$

而方差的正平方根，则被称为标准差，记为$\sigma(X)$。

方差与标准差的功能相似，它们都是用来描述随机变量取值的集中与分散程度(即散布大小)的两个特征数。方差与标准差愈小，随机变量的取值愈集中；方差与标准差愈大，随机变量的取值愈分散。

方差与标准差之间的差别主要在量纲上，由于标准差与所讨论的随机变量、数学期望有相同的量纲，其加减$E(X)\pm k\sigma(X)$是有意义的(k为正实数)，所以在实际中，人们比较乐意选用标准差，但标准差的计算必须通过方差才能算得。

另外要指出的是：如果随机变量X的数学期望存在，其方差不一定存在；而当X的方差存在时，则$E(X)$必定存在，其原因在于$|x|\leq x^2+1$总是成立的。

关于方差的计算，我们这里不再举例，主要是要了解方差的性质：

- 最重要的性质： $Var(X)=E(X^2)-[E(X)]^2$ 。
- 常数的方差为0，即 $Var(c)=0$ ，其中 c 是常数。
- 若 a,b 是常数，则 $Var(aX+b)=a^2Var(X)$ 。
- 若随机变量 X 与 Y 相互独立，则有 $Var(X\pm Y)=Var(X)+Var(Y)$ 。

2.4. 常见离散分布

接下来我们来了解以下常见的离散分布，以及对应的数学期望和方差

2.4.1. 二项分布

首先我们了解一个概念，伯努利试验。伯努利试验是在同样的条件下重复地、相互独立地进行的一种随机试验，其特点是该随机试验只有两种可能结果：发生或者不发生。我们假设该项试验独立重复地进行了n次，那么就称这一系列重复独立的随机试验为n重伯努利试验。

如果记X为n重伯努利试验中成功(记为事件A)的次数，则X的可能取值为$0,1,\cdots,n$。记p为每次试验A发生的概率，则有$p(A)=p,p(\bar{A})=1-p$。下面我们求X的分布列，即求事件$\{X=k\}$的概率，这意味着成功了k次，失败了$n-k$次，从组合可以得到，$P(X=k)=C_n^kp^k(1-p)^{n-k}$, $k=0,1,\cdots,n$。这个分布就被称为二项分布，记为$X\sim b(n,p)$。

下面我们来求一下二项分布的期望和方差：

$$
\begin{aligned}
E(X)&=\sum_{k=0}^{n}kC_n^kp^k(1-p)^{n-k}\\
&=\sum_{k=1}^{n}npC_{n-1}^{k-1}p^{k-1}(1-p)^{(n-1)-(k-1)}\\
&=np\sum_{k=1}^{n}C_{n-1}^{k-1}p^{k-1}(1-p)^{(n-1)-(k-1)}\\
&=np[p+(1-p)]^{n-1}=np\\
\text{其中，}C_n^k&=\frac{n!}{(n-k)!k!}=\frac{n\cdot(n-1)!}{((n-1)-(k-1))!(k\cdot(k-1)!)}\\
&=\frac{n}{k}C_{n-1}^{k-1}
\end{aligned}
$$

而

$$
\begin{aligned}
E(X^2) &= \sum_{k=0}^{n} k^2 C_n^k p^k (1-p)^{n-k} = \sum_{k=0}^{n} (k-1+1) k C_n^k p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} k(k-1) C_n^k p^k (1-p)^{n-k} + \sum_{k=0}^{n} k C_n^k p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} k(k-1) C_n^k p^k (1-p)^{n-k} + np \\
&= \sum_{k=0}^{n} k(k-1) C_n^k p^k (1-p)^{n-k} + np \\
&= n \cdot (n-1) \sum_{k=2}^{n} p^2 C_{n-2}^{k-2} p^{k-2} (1-p)^{(n-2)-(k-2)} + np \\
&= n(n-1)p^2 + np \\
\text{其中，} C_n^k &= \frac{n!}{(n-k)!k!} = \frac{n \cdot (n-1) \cdot (n-2)!}{((n-2)-(k-2))!(k \cdot (k-1) \cdot (k-2)!)} \\
&= \frac{n \cdot (n-1)}{k \cdot (k-1)} C_{n-2}^{k-2}
\end{aligned}
$$

因此，$Var(X) = E(X^2) - [E(X)]^2 = n(n-1)p^2 + np - n^2p^2 = np(1-p)$

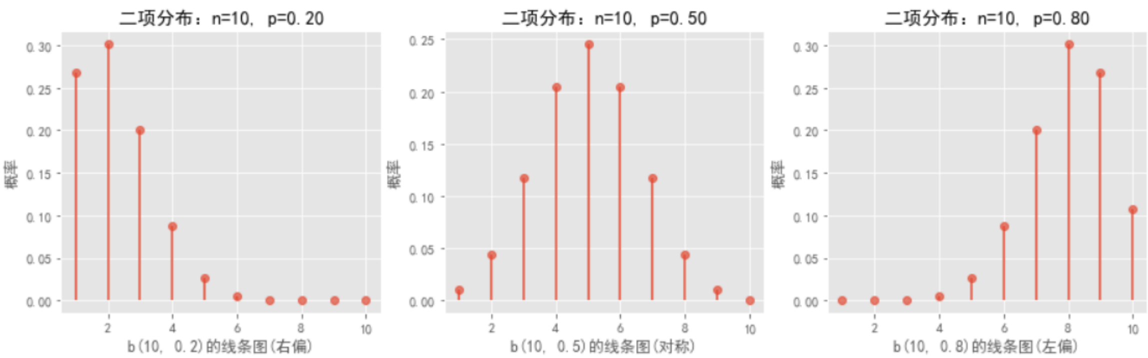
当$n=1$时，我们可以看到此时二项分布为$b(1,p)$，此时就是$0-1$分布。

我们通过图像来简单看一下二项分布

```python
# 使用scipy的pmf和cdf画图
from scipy.stats import binom

n=10
x=np.arange(1,n+1,1)

fig, axes = plt.subplots(1, 3, figsize=(15, 4))
for i, (p, line_type) in enumerate(zip([0.2, 0.5, 0.8], ['右偏', '对称', '左偏'])):
    pList=binom.pmf(x,n,p)
    axes[i].plot(x,pList,marker='o',alpha=0.7,linestyle='None')
    axes[i].vlines(x, 0, pList)
    axes[i].set_xlabel(f'b({n}, {p})的线条图({line_type})')
    axes[i].set_ylabel('概率')
    axes[i].set_title(f'二项分布: n={n}, p={p:.2f}')
plt.show()
```



2.4.2. 泊松分布

泊松分布的概率分布列是$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$, $k=0,1,2,\cdots$，其中参数$\lambda>0$，记为$X\sim P(\lambda)$。泊松分布是一种常用的离散分布, 它常与单位时间 (或单位面积、单位产品等)上 的计数过程相联系, 譬如,

- 在一天内，来到某商场的顾客数。（λ 就是单位时间内商场的顾客数）
- 在单位时间内，一电路受到外界电磁波的冲击次数。
- 1平方米内，玻璃上的气泡数。
- 一铸件上的砂眼数。
- 在一定时期内，某种放射性物质放射出来的 α -粒子数，等等。

下面我们来看以下泊松分布的期望和方差。

$$
\begin{aligned}
E(X) &= \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda e^{-\lambda}e^{\lambda} = \lambda \\
\text{其中，} e(\lambda) &= 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots \\
&= \sum_{k=0}^{\infty}\frac{x^k}{k!} = \sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} \\
E(X^2) &= \sum_{k=0}^{\infty} k^2\frac{\lambda^k}{k!}e^{-\lambda} = \sum_{k=1}^{\infty} k\frac{\lambda^k}{(k-1)!}e^{-\lambda} \\
&= \sum_{k=1}^{\infty}(k-1+1)\frac{\lambda^k}{(k-1)!}e^{-\lambda} \\
&= \sum_{k=1}^{\infty}(k-1)\frac{\lambda^k}{(k-1)!}e^{-\lambda} + \lambda \\
&= \lambda^2\sum_{k=1}^{\infty}(k-2)\frac{\lambda^{k-2}}{(k-2)!}e^{-\lambda} + \lambda \\
&= \lambda^2 + \lambda \\
Var(X) &= E(X^2) - [E(X)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda
\end{aligned}
$$

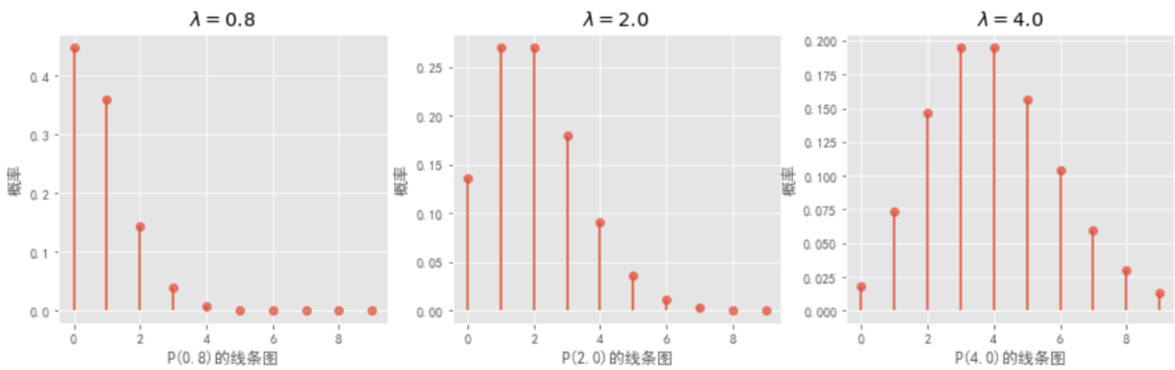
我们再来看一下泊松分布的图像

```python
import math

def poisson(lamb, k):
    return math.exp(-lamb) * lamb ** k / math.factorial(k)

x = [i for i in range(10)]
lambs = [0.8, 2.0, 4.0]
ps = [[poisson(lamb, i) for i in range(10)] for lamb in lambs]

fig, axes = plt.subplots(1, 3, figsize=(15, 4))
for i, (lamb, p) in enumerate(zip(lambs, ps)):
    axes[i].plot(x, p,marker='o',alpha=0.7,linestyle='None')
    axes[i].vlines(x, 0, p)
    axes[i].set_xlabel(f'P({lamb})的线条图')
    axes[i].set_ylabel('概率')
    axes[i].set_title(f'$\lambda={lamb}$')
```



2.5. 常见连续分布

2.5.1. 正态分布

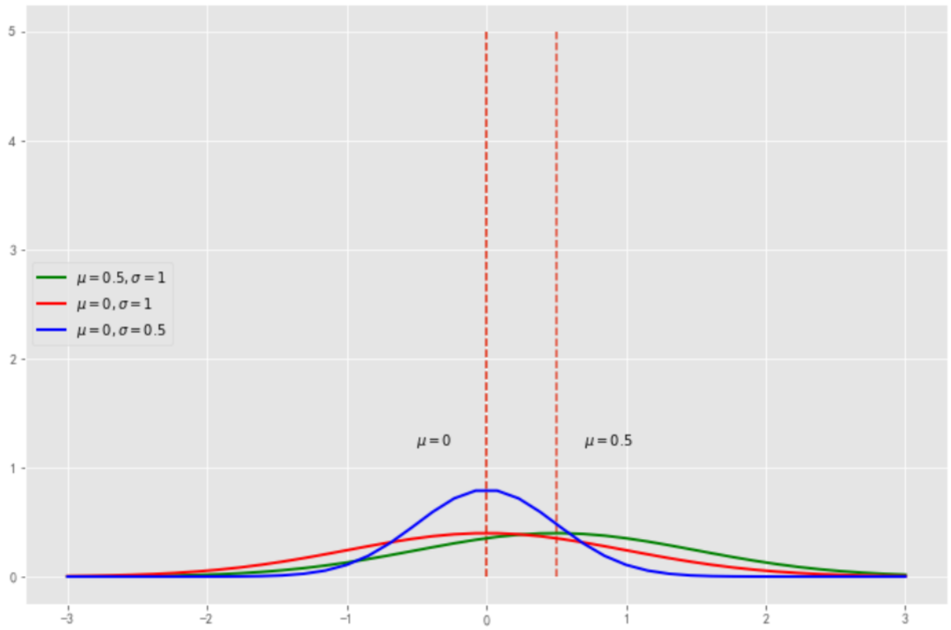
若随机变量X的密度函数为$p(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$，$-\infty<x<+\infty$，则称$X$服从正态分布，记作$X\sim N(\mu,\sigma^2)$。我们来看一下正态分布的图像：

```python
def normal(mu, sigma):
    return np.exp(-0.5 * (x-  mu) ** 2 / (sigma ** 2)) / (math.sqrt(2 * math.pi) * sigma)

x = np.linspace(-3, 3, 40)
params = [(0.5, 1), (0, 1), (0, 0.5)]
colors = ["green", "red", "blue"]
```

```python

plt.figure(figsize=(12, 8))
for i, (param, color) in enumerate(zip(params, colors)):
    mu, sigma = param
    y = normal(mu, sigma)
    plt.plot(x, y, color=color, linewidth=2, label=f"$\mu={mu},
\sigma={sigma}$")
    plt.vlines(mu, 0, 5, linestyles='--')
plt.text(-0.5, 1.2, "$\mu=0$")
plt.text(0.7, 1.2, "$\mu=0.5$")
plt.legend(loc="center left")
plt.show()
```



从图中可以看出：如果固定σ，改变μ的的值，则图形沿x轴平移，而不改变其形状。也就是说正态密度函数的位置由参数μ所确定，因此亦称μ为位置参数；同时如果固定μ，改变σ的的值，则分布的位置不变，但σ愈小，曲线呈高而瘦，分布较为集中；σ愈大，曲线呈矮而胖，分布较为分散。也就是说正态密度函数的尺度由参数σ所确定，因此称σ为尺度参数。

我们将$\mu=0,\sigma=1$的正态分布称为标准正态分布。对于标准正态分布，有

- $\Phi(-u)=1-\Phi(-u)$
- $P(U>u)=1-\Phi(u)$
- $P(a<U<b)=\Phi(b)-\Phi(a)$
- $P(|U|<c)=2\Phi(c)-1(c\geq 0)$

我们再来看一下正态分布的数学期望和方差，由于$U=(X-\mu)/\sigma\sim N(0,1)$

$$E(U)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}ue^{-\frac{u^2}{2}}du$$

因为被积函数是奇函数，所以$E(U)=0, E(X)=E(\mu+\sigma U)=\mu$

$$
\begin{aligned}
Var(U) = E(U^2)-[E(U)]^2 = E(U^2) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}u^2e^{-\frac{u^2}{2}}du \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}ud(-e^{-\frac{u^2}{2}}) = \frac{1}{\sqrt{2\pi}}\left(-ue^{-\frac{u^2}{2}}+\int_{-\infty}^{\infty}e^{-\frac{u^2}{2}}du\right) \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{u^2}{2}}du \\
&= \frac{1}{\sqrt{2\pi}}\cdot\sqrt{2\pi}=1
\end{aligned}
$$

$$
\begin{aligned}
\text{其中，}\int_{-\infty}^{\infty}e^{-\frac{u^2}{2}}du &= \sqrt{\int_{-\infty}^{\infty}e^{-\frac{x^2}{2}}dx\int_{-\infty}^{\infty}e^{-\frac{y^2}{2}}dy} \\
&= \sqrt{\iint e^{-\frac{x^2+y^2}{2}}dxdy} = \sqrt{\int_0^{2\pi}d\theta\int_0^{+\infty}re^{-\frac{r^2}{2}}dr} \\
&= \sqrt{2\pi\cdot(-e^{-\frac{r^2}{2}})|_0^{+\infty}} = \sqrt{2\pi}
\end{aligned}
$$

所以$Var(X)=Var(\mu+\sigma U)=\sigma^2\cdot Var(U)=\sigma^2$。在正态分布中，有一个$3\sigma$原则需要注意：

$$
\begin{gathered}
P(\mu-k\sigma<X<\mu+k\sigma)=P(|\frac{X-\mu}{\sigma}|<k)=2\Phi(k)-1 \\
P(\mu-\sigma<X<\mu+\sigma)=2\Phi(1)-1=0.6826 \\
P(\mu-2k\sigma<X<\mu+2k\sigma)=2\Phi(2)-1=0.9545 \\
P(\mu-3k\sigma<X<\mu+3k\sigma)=2\Phi(3)-1=0.9973
\end{gathered}
$$

2.5.2. 均匀分布

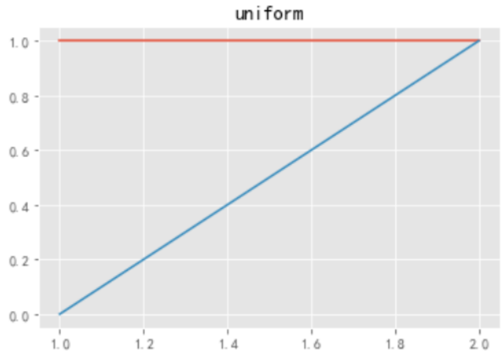
其实我们前面在讲期望的时候就已经说过均匀分布了，其概率密度函数为：

$$
p(x)=\begin{cases}\dfrac{1}{b-a}, & a\le x<b \\ 0, & \text{其他}\end{cases}
$$

其图像如下图所示：

```
a, b = 1.0, 2.0
x = np.linspace(a, b)
y = np.full_like(x, fill_value=1 / (b - a))

plt.plot(x, y)
plt.plot(x, (x - a) / (b - a))
plt.title("uniform");
```



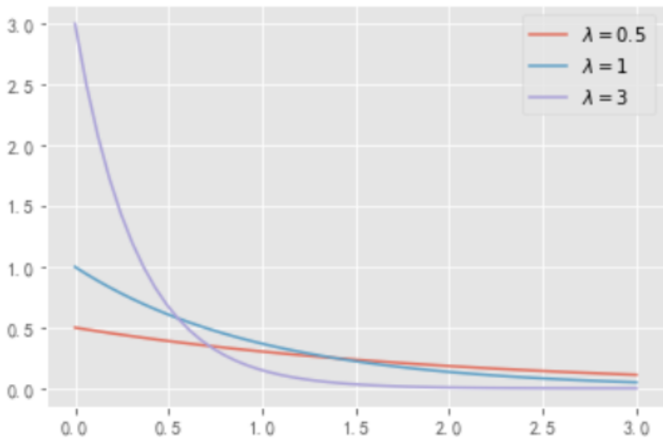
$$E(X) = \int_a^b \frac{x}{b-a} dx = \frac{x^2}{2(b-a)} \Big|_a^b = \frac{a+b}{2}$$
$$E(X^2) = \int_a^b \frac{x^2}{b-a} dx = \frac{x^3}{3(b-a)} \Big|_a^b = \frac{a^2+ab+b^2}{3}$$
$$Var(X) = E(X^2) - [E(X)]^2 = \frac{(b-a)^2}{12}$$

2.5.3. 指数分布

若随机变量X的密度函数为

$$p(x) = \begin{cases} \lambda e^{-\lambda}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$
$$F(x) = \begin{cases} 1 - e^{-\lambda}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

指数分布是一种偏态分布，由于指数分布随机变量只可能取非负实数,所以指数分布常被用作各种“寿命”分布,譬如电子元器件的寿命、动物的寿命、电话的通话时间、随机服务系统中的等待时间等都可假定服从指数分布.指数分布在可靠性与排队论中有着广泛的应用。我们来看一下其分布：



最后再来看一下指数分布的期望和方差：

$$
\begin{aligned}
E(X) &= \int_0^{+\infty} x\lambda e^{-\lambda x}dx = \int_0^{+\infty} xd(-e^{-\lambda x}) \\
&= -xe^{-\lambda x}\Big|_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x}dx \\
&= -\frac{1}{\lambda}\cdot e^{-\lambda x}\Big|_0^{+\infty} = \frac{1}{\lambda} \\
E(X^2) &= \int_0^{+\infty} x^2\lambda e^{-\lambda x}dx = \int_0^{+\infty} x^2 d(-e^{-\lambda x}) \\
&= -x^2e^{-\lambda x}\Big|_0^{+\infty} + 2\int_0^{+\infty} xe^{-\lambda x}dx \\
&= \frac{2}{\lambda}E(X) = \frac{2}{\lambda^2} \\
Var(X) &= E(X^2) - [E(X)]^2 = \frac{1}{\lambda^2}
\end{aligned}
$$

指数分布有一项特殊的性质——无记忆性。如果随机变量$X \sim Exp(\lambda)$，那么对于任意$s>0, t>0$，有$P(X>s+t|X>s) = P(X>t)$，证明如下：

$$
\{X>s+t\} \subset \{X>s\}
$$

$$
P(X>s+t|X>s) = \frac{P(X>s+t)}{P(X>s)} = \frac{1-(1-e^{-(s+t)})}{1-(1-e^{-s})} = e^{-t} = P(X>t)
$$

2.7. 分布的其他特征数

除了数学期望和方差外，随机变量还有一些很重要的特征数，下面我们逐一给出定义和解释。

2.7.1. k阶矩

设X为随机变量，k为正整数，如果$\mu_k = E(X^k)$都存在，则称μ_k为X的k阶原点矩；如果$v_k = E(X-E(X))^k$为X的k阶中心矩。我们举个简单例子：

设随机变量$X \sim N(0, \sigma^2)$，则

$$
\begin{aligned}
\mu_k = E(X^k) &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} x^k exp\{-\frac{x^2}{2\sigma^2}\}dx \\
&== \frac{\sigma^k}{\sqrt{2\pi}}\int_{-\infty}^{\infty} u^k exp\{-\frac{u^2}{2}\}du
\end{aligned}
$$

其中$, u = \dfrac{x}{\sigma}$。

当k为奇数时，上述被积函数是基函数，故$\mu_k = 0$，$k = 1, 3, 5, \cdots$。当k为偶数时，上述被积函数是偶函数，再利用变换$z = u^2/2$，可得

$$
\begin{aligned}
u_k &= \sqrt{\frac{2}{\pi}}\sigma^k 2^{\frac{k-1}{2}} \int_0^{\infty} z^{(k-1)/2} e^{-z} dz \\
&= \sqrt{\frac{2}{\pi}}\sigma^k 2^{(k-1)/2}\Gamma(\frac{k+1}{2}) \\
&= \sigma^k (k-1)(k-3)\cdots 1 \quad k = 2, 4, 6, \cdots
\end{aligned}
$$

故$N(0, \sigma^2)$分布的前四阶原点矩为$\mu_1 = 0, \quad \mu_2 = \sigma^2, \quad \mu_3 = 0, \quad \mu_4 = 3\sigma^4$，又因为$E(X) = 0$，所以原点矩等于中心矩。

2.7.2. 变异系数

方差(或标准差)反映了随机变量取值的波动程度，但在比较两个随机变量的波动大小时，如果仅看方差(或标准差)的大小有时会产生不合理的现象。这有两个原因：(1)随机变量的取值有量纲，不同量纲的随机变量用其方差(或标准差)去比较它们的波动大小不太合理(2)在取值的量纲相同的情况下，取值的大小有一个相对性问题，取值较大的随机变量的方差(或标准差)也允许大一些。所以要比较两个随机变量的波动大小时，在有些场合使用以下定义的变异系数来进行比较，更具可比性。

设随机变量X的二阶矩存在，则称比值$C_v(X) = \dfrac{\sqrt{Var(X)}}{E(X)} = \dfrac{\sigma(X)}{E(X)}$为$X$的变异系数。变异系数是一个无量纲的量，从而可以消除量纲对波动的影响。

2.7.3. 分位数

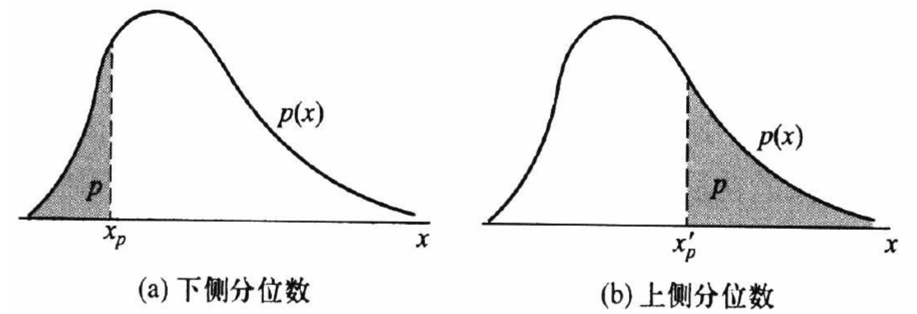
设连续随机变量X的分布函数为$F(x)$，密度函数为$p(x)$，对任意的$p \in (0, 1)$，称满足条件

$$
F(x_p) = \int_{-\infty}^{x_p} p(x)dx = p
$$

的x_p为此分布的p分位数，又称下侧p分位数。分位数x_p把密度函数下的面积分为两块，左侧面积恰好为p。同理我们称满足条件

$$
1 - F(x'_p) = \int_{x'_p}^{\infty} p(x)dx = p
$$

的x'_p为此分布的上侧p分位数。



(a) 下侧分位数　(b) 上侧分位数

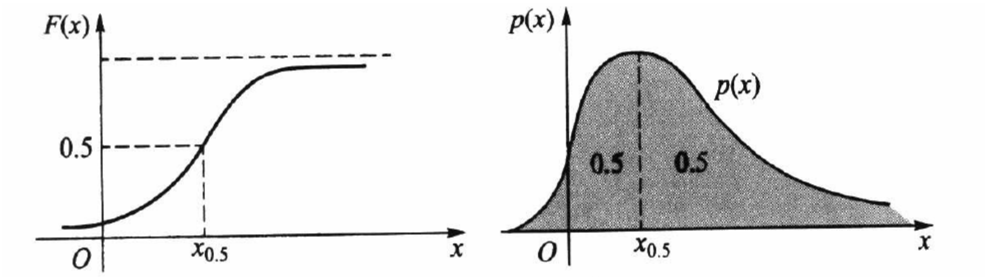
分位数与上侧分位数是可以相互转换的，其转换公式如下：

$$x'_p = x_{1-p}, \quad x_p = x'_{1-p}$$

2.7.4. 中位数

设连续随机变量X的分布函数为$F(x)$，密度函数为$p(x)$，称$p=0.5$时的p分位数$x_{0.5}$为此分布的中位数，即

$$F(x_p) = \int_{-\infty}^{x_{0.5}} p(x)dx = 0.5$$



中位数与均值一样都是随机变量位置的特征数，但在某些情况可能中位数比均值更能说明问题。比如在统计某个行业工资水平，直接使用平均数可能会导致大部分人被平均，这种情况用中位数就更能反映工资水平。

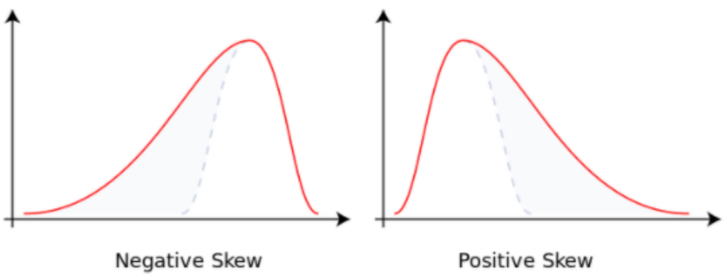
我们使用`python`来计算分位数水平。

```
from scipy.stats import norm
print("标准正态分布的0.25分位数：",norm(loc=0,scale=1).ppf(0.25))   # 使用ppf计算分位数点
print("标准正态分布的0.5分位数：",norm(loc=0,scale=1).ppf(0.5))
print("标准正态分布的0.75分位数：",norm(loc=0,scale=1).ppf(0.75))
print("标准正态分布的0.95分位数：",norm(loc=0,scale=1).ppf(0.95))
```

```
标准正态分布的0.25分位数： -0.6744897501960817
标准正态分布的0.5分位数： 0.0
标准正态分布的0.75分位数： 0.6744897501960817
标准正态分布的0.95分位数： 1.6448536269514722
```

2.7.5. 偏度系数

设随机变量X的前三阶矩存在，则比值$\beta_s = \frac{v_3}{v_2^{3/2}} = \frac{E(X-E(X))^3}{[Var(X)]^{3/2}}$称为$X$的偏度系数，简称偏度。当$\beta_s > 0$，则该分布为右偏；当$\beta_s < 0$，则该分布为左偏。偏度是描述分布偏离对称性程度的一个特征数，分布的三阶中心矩决定偏度的符号，而分布的标准差$\sigma(X)$决定偏度大小。



2.7.6. 峰度系数

设随机变量X的前四阶矩存在，则比值$\beta_k = \frac{v_4}{v_2^2} - 3 = \frac{E(X-E(X))^4}{[Var(X)]^2} - 3$称为$X$的峰度系数，简称峰度。峰度是描述分布尖峭程度和尾部粗细的一个特征数。

3. 多维随机变量及其分布

在某些随机现象中，对每个样本点ω只用一个随机变量去描述是不够的，这时可能需要两个以上随机变量，我们先来研究联合分布函数，再研究离散随机变量的联合分布列、连续随机变量的联合密度函数等。

3.1. 多维随机变量及其联合分布

对于任意的n个实数$x_1, x_2, \cdots, x_n$，n个事件$\{X_1 \le x_1\}, \{X_2 \le x_2\}, \cdots, \{X_n \le x_n\}$同时发生的概率$F(x_1, x_2, \cdots, x_n) = P(X_1 \le x_1, X_2 \le x_2, \cdots, X_n \le x_n)$称为$n$维随机变量$(X_1, X_2, \cdots, X_n)$的联合分布函数。

3.1.1. 联合分布列

如果二维随机变量(X,Y)只取有限个数对(x_i,y_i)，则称(X,Y)为二维离散随机变量，称$P_{ij}=P(X=x_i,Y=y_j)$，　$i,j=1,2,\cdots$为(X,Y)的联合分布列，也可以用如下表格表示联合分布列：

X	Y				
	y_1	y_2	$\cdots$	y_j	$\cdots$
x_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$
x_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	

求二维离散随机变量的联合分布列，关键是要写出二维随机变量可能取得数对及其发生得概率，我们举一个简单例子来说明。

从 $1,2,3,4$ 中任取一数记为 X， 再从 $1,2,\cdots,X$ 中任取一数记为 Y。 求 (X,Y) 的联合分布列及 $P(X=Y)$。

分析：(X,Y)为二维离散随机变量，其中X的分布列为
$P(X=i)=1/4,\quad i=1,2,3,4$。$Y$的可能取值也是$1,2,3,4$，若记$j$为$Y$的取值，则当$j>i$时，有$P(X=i,Y=j)=0$，当$1\le j\le i\le 4$时，由乘法公式：
$P(X=i,Y=j)=P(X=i)P(Y=j|X=i)=\dfrac{1}{4}\times\dfrac{1}{i}$。

因此(X,Y)的联合分布列为

X	Y			
	1	2	3	4
1	1/4	0	0	0
2	1/8	1/8	0	0
3	1/12	1/12	1/12	0
4	1/16	1/16	1/16	1/16

因此，$P(X=Y)=p_{11}+p_{22}+p_{33}+p_{44}=\dfrac{1}{4}+\dfrac{1}{8}+\dfrac{1}{12}+\dfrac{1}{16}=\dfrac{25}{48}$。

3.1.2. 联合密度函数

如果存在二元非负函数$p(x,y)$，使得二维随机变量(X,Y)的分布函数$F(x,y)$可表示为：

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}p(u,v)dudv$$

则称(X,Y)为二维连续随机变量，$p(x,y)$为(X,Y)的联合密度函数。在$F(x,y)$偏导数存在的点上有$p(x,y)=\dfrac{\partial^2}{\partial x\partial y}F(x,y)$。

同样的，我们举一个简单的例子。设(X,Y)的联合密度函数为

$$p(x,y)=\begin{cases}6e^{-2x-3y}, & x>0,y>0\\ 0,其他\end{cases}$$

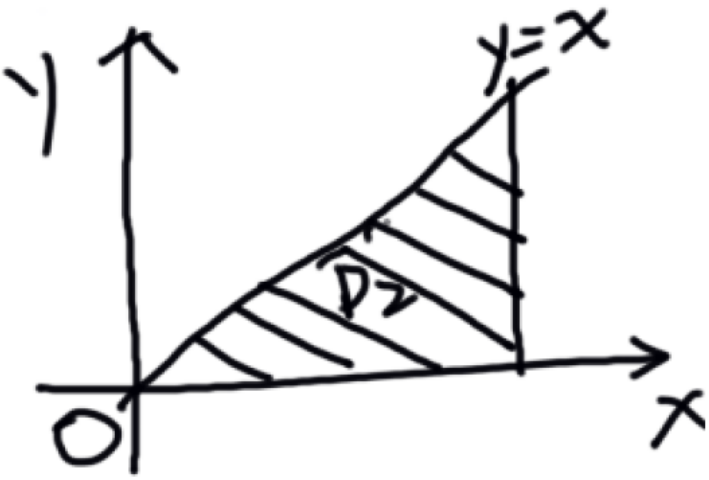
试求(1)$P(X<1,Y>1)$；(2)$P(X>Y)$

解析：（1）对应积分区域如下所示：

$$P(X<1,Y>1)=P((X,Y)\in D_1)=\int_0^1\int_1^{+\infty}6e^{-2x-3y}dxdy$$
$$=(1-e^{-2})e^{-3}$$

(2)对应积分区域如下所示：

$$P(X>Y)=P((X,Y)\in D_2)=\int_0^{\infty}\int_0^{x}6e^{-2x-3y}dxdy$$
$$=\left(-e^{-2x}+\frac{2}{5}e^{-5x}\right)\Big|_0^{+\infty}=\frac{3}{5}$$



3.2. 边际分布与随机变量的独立性

二维联合分布函数含有丰富的信息，主要包括：

- 每个分量的分布(每个分量的所有信息)，即边际分布
- 两个分量之间的关联程度，可用协方差和相关系数来描述
- 给定一个分量时，求另一个分量的分布，即条件分布

3.2.1. 边际分布函数

如果在二维随机变量 (X,Y) 的联合分布函数 $F(x,y)$ 中令 $y\to\infty$， 由于 $\{Y<\infty\}$ 为必然事件， 故可得

$$\lim_{y\to\infty}F(x,y)=P(X\leqslant x,Y<\infty)=P(X\leqslant x)$$

这是由 (X,Y) 的联合分布函数 $F(x,y)$ 求得的 X 的分布函数， 被称为 X 的边际分布, 记为$F_X(x)=F(x,\infty)$。类似地， 在 $F(x,y)$ 中令 $x\to\infty$， 可得 Y 的边际分布 $F_Y(y)=F(\infty,y)$。我们这里举一个简单例子：

设二维随机变量 (X,Y) 的联合分布函数为

$$F(x,y)=\begin{cases}1-e^{-x}-e^{-y}+e^{-x-y-\lambda xy}, & x>0,y>0.\\ 0, & \text{其他}\end{cases}$$

这个分布被称为二维指数分布，其中参数 $\lambda>0$。

由此联合分布函数$F(x,y)$，容易获得X与Y的边际分布函数为

$$F_X(x)=F(x,\infty)=\begin{cases}1-e^{-x}, & x>0\\ 0, & x\le 0\end{cases}$$
$$F_Y(y)=F(\infty,y)=\begin{cases}1-e^{y}, & y>0\\ 0, & y\le 0\end{cases}$$

不同的λ对应不同的二维指数分布，但是它们的边际分布与λ无关。这说明二维联合分布不仅含有每个分量的概率分布，而且还含有两个变量X与Y间关系的信息。

3.2.2. 边际分布列

在二维离散随机变量(X,Y)的联合分布列$\{P(X=x,Y=y\}$中，对j求和所得的分布列

$$\sum_{j=1}^{\infty} P(X=x, Y=y_j) = P(X=x_i), \quad i=1,2,\cdots$$

被称为X的分布列。类似的，对i求和所得的分布列

$$\sum_{i=1}^{\infty} P(X=x_i, Y=y_j) = P(Y=y_i), \quad j=1,2,\cdots$$

被称为Y的分布列。

我们举个简单的例子。设二维随机变量(X,Y)有如下的联合分布列

X	Y		
	1	2	3
0	0.09	0.21	0.24
1	0.07	0.12	0.27

求X与Y的分布列。

解析：对上面联合分布中，每行求和得0.54和0.46，这就是X对应的概率；再对每一列求和，分别得到$0.16, 0.33, 0.51$，这就是Y对应的概率。我们将其写在联合分布列，则有：

X	Y			P(X=i)
	1	2	3	
0	0.09	0.21	0.24	0.54
1	0.07	0.12	0.27	0.46
P(Y=j)	0.16	0.33	0.51	1

3.2.3. 边际密度函数

如果二维连续随机变量(X,Y)的联合概率密度函数为$p(x,y)$，因为

$$F_X(x) = F(x,\infty) = \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} p(u,v)dv\right) du = \int_{-\infty}^{x} p_X(u)du$$

$$F_Y(x) = F(\infty,y) = \int_{-\infty}^{y} \left(\int_{-\infty}^{\infty} p(u,v)du\right) dv = \int_{-\infty}^{y} p_Y(v)dv$$

其中,

$$p_X(x) = \int_{-\infty}^{\infty} p(x,y)dy$$

$$p_Y(y) = \int_{-\infty}^{\infty} p(x,y)dx$$

上式所给的$p_X(x)$为X的边际密度函数，$p_Y(y)$为Y的边际密度函数。

我们来看下面这个例子。设二维随机变量(X,Y)的概率密度函数为

$$p(x,y)=\begin{cases}1, & 0<x<1,|y|<1\\ 0, & \text{其他}\end{cases}$$

试求：边际概率密度函数$p_X(x)$和$p_Y(y)$

分析：首先我们可以看到$p(x,y)$的非零区域。

对于$p_X(x)$，当$x\le 0$或$x\ge 1$时，有$p_X(x)=0$，而当$0<x<1$时，有

$$p_X(x)=\int_{-x}^{x}dy=2x$$

所以X的边际密度函数为

$$p_X(x)=\begin{cases}2x, & 0<x<1\\ 0, & \text{其他}\end{cases}$$

再求$p_Y(y)$。

当$y\le -1$，或$y\ge 1$时，有$p_Y(y)=0$。而当$-1<y<0$时，有

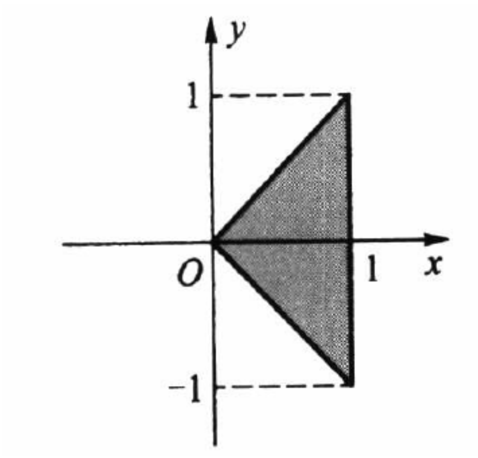
$$p_Y(y)=\int_{-y}^{1}dx=1+y$$

而当$0<y<1$时，有

$$p_Y(y)=\int_{y}^{1}dx=1-y$$

所以Y的边际密度函数为

$$p_Y(y)=\begin{cases}1+y, & -1<y<0\\ 1-y, & 0<y<1\\ 0, & \text{其他}\end{cases}$$



3.2.4. 随机变量间的独立性

设n维随机变量$(X_1,X_2,\cdots,X_n)$的联合分布函数$F(x_1,x_2,\cdots,x_n)$，$F_i(x_i)$为X_i的边际分布函数。如果对任意n个实数$x_1,x_2,\cdots,x_n$，有$F(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}F_i(x_i)$，则称$X_1,X_2,\cdots,X_n$相互独立。

对于离散随机变量，如果对其任意n个取值$x_1,x_2,\cdots,x_n$，有$P(X_1=x_1,X_2=x_2,\cdots,X_n=x_n)=\prod_{i=1}^{n}P(X_i=x_i)$，则称$X_1,X_2,\cdots,X_n$相互独立。

对于连续随机变量，如果对任意n个实数$x_1,x_2,\cdots,x_n$，有$p(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}p_i(x_i)$，则称$X_1,X_2,\cdots,X_n$相互独立。

3.3. 多维随机变量的特征数

3.3.1. 数学期望

若二维随机变量(X,Y)的分布用联合分布列$P(X=x,Y=y)$或用联合密度函数$p(x,y)$表示，则$Z=g(X,Y)$的数学期望为

$$E(Z)=\begin{cases}\sum_i\sum_j g(x_i,y_j)P(X=x_i,Y=y_j), & \text{离散}\\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g(x,y)p(x,y)dxdy, & \text{连续}\end{cases}$$

还要指出，在连续(离散)有：

- 当 $g(X,Y)=X$ 时，可得 X 的数学期望为

$$E(X)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xp(x,y)dxdy=\int_{-\infty}^{\infty}xp_X(x)dx$$

- 当 $g(X,Y)=(X-E(X))^2$ ，可得 X 的方差为

$$\begin{aligned}Var(X)=E(X-E(X))^2&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-E(X))^2p(x,y)dxdy\\&=\int_{-\infty}^{\infty}(x-E(X))^2p_X(x)dx\end{aligned}$$

类似的，还可以给出Y的数学期望与反差的公式。

- $E(X+Y)=E(X)+E(Y)$
- 若随机变量 X 与 Y 相互独立，则有 $E(XY)=E(X)E(Y)$
- 若随机变量 X 与 Y 相互独立，则有 $Var(X\pm Y)=Var(X)+Var(Y)$

来个简单的例子。设一袋中装有m个颜色各不相同的球，每次从中任取一个，有放回的摸取n次，以X表示在n次磨球中摸到球的不同颜色的数目，求$E(X)$。

分析：直接写出X的分布列较为困难，其原因在于：若第i种颜色的球被取到过，则此种颜色的球又可被取到过一次、二次......n次，情况较多，而其对立事件“第i种颜色的球没被取到过”的概率容易写出为P(第i种颜色的球在n次摸球中一次也没被摸到)=$(1-\frac{1}{m}^{n})$。令

$$X_i=\begin{cases}1, & \text{第}i\text{种颜色的球在}n\text{次摸球中至少被摸到一次，}\\ 0, & \text{第}i\text{种颜色的球在}n\text{次摸球中一次也没被摸到，}\end{cases}\qquad i=1,2,\cdots,m$$

这些X_i相当于是计数器，分别记录下第i种颜色的球是否被取到过，X是取到过的不同颜色的总数，所以$X=\sum_{i=1}^{m}X_i$，由$P(X_i=0)=(1-\frac{1}{m})^n$。可得$E(X_i)=P(X_i=1)=1-(1-\frac{1}{m})^n$。所以$E(X)=mE(X_i)=m\left[1-(1-\frac{1}{m})^n\right]$。

3.3.2. 协方差

二维联合分布中除含各分量的边际分布外，还含有两个分量间相互关系的信息，描述这种相互关联程度的一个特征数就是协方差。定义如下：设(X,Y)是一个二维随机变量，若$E[(X-E(X))(Y-E(Y))]$存在，则称此数学期望为X与Y的协方差，或称为X与Y的相关矩，并记为$Cov(X,Y)=E[(X-E(X))(Y-E(Y))]$，特别的，有$Cov(X,X)=Var(X)$。

- 当 $Cov(X,Y)>0$ 时，称 X 与 Y 正相关，这时两个偏差 $(X-E(X))$ 与 $(Y-E(Y))$ 有同时增加或同时减少的倾向。由于 $E(X)$，$E(Y)$ 为常数，因此可以等价于 X 与 Y 有同时增加或减少的倾向，这就是正相关。
- 当 $Cov(X,Y)<0$ 时，称 X 与 Y 负相关，这时这就是有 X 增加而 Y 减少的倾向，或有 Y 增加而 X 减少的倾向，这就是负相关。
- 当 $Cov(X,Y)=0$ 时，称 X 与 Y 不相关，注意不能等同于 X 与 Y 独立。

我们来看一个简单例子，设二维随机变量(X,Y)的联合密度函数为

$$p(x,y)=\begin{cases}3x, & 0<y<x<1,\\ 0, & 其他\end{cases}$$

试求$Cov(X,Y)$。

分析：

$$\begin{aligned}E(X)&=\int_0^1\int_0^x x\cdot 3xdydx=\frac{3}{4}\\ E(Y)&=\int_0^1\int_0^x y\cdot 3xdydx=\frac{3}{8}\\ E(XY)&=\int_0^1\int_0^x xy\cdot 3xdydx=\frac{3}{10}\\ Cov(X,Y)&=E(XY)-E(X)E(Y)=\frac{3}{10}-\frac{3}{4}\times\frac{3}{8}=\frac{3}{160}>0\end{aligned}$$

因此X与Y不相互独立。

3.3.3. 相关系数

正如方差与协方差一样，因为协方差是有量纲的量，所以引入新的概念——相关系数：设(X,Y)是一个二维随机变量，且$Var(X)=\sigma_X^2>0, Var(Y)=\sigma_Y^2>0$，则

$$Corr(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}$$

这里我们引入一个施瓦茨不等式：对任意二维随机变量(X,Y)，若X与Y的方差都存在，且记$\sigma_X^2=Var(X), \sigma_Y^2=Var(Y)$，则有$[Cov(X,Y)]^2\le\sigma_X^2\sigma_Y^2$。证明如下：

不妨设$\sigma_X^2>0$，因为当$\sigma_X^2=0$时，说明X为常数，其与Y协方差为零，结论成立。若$\sigma_X^2>0$，考虑如下二次函数：

$$g(t)=E[t(X-E(X))+(Y-E(Y))]^2=t^2\sigma_X^2+2t\cdot Cov(X,Y)+\sigma_Y^2$$

由于上面的二次函数非负，而平方项系数σ_X^2为正，所以判别式$\Delta\le 0$，即$[2Cov(X,Y)]^2-4\sigma_X^2\sigma_Y^2\le 0$，移项后可得施瓦茨不等式。从该不等式可以得到$-1\le Corr(X,Y)\le 1$。

- $Corr(X,Y)=\pm 1$ 的充要条件是 X 与 Y 间几乎处处有线性关系, 即存 在 $a(\neq 0)$ 与 b， 使得 $P(Y=aX+b)=1$ 。
- 相关系数 $Corr(X,Y)$ 刻画了 X 与 Y 之间的线性关系强弱， 因此也常称其为“线性相关系数”。
- 若 $Corr(X,Y)=0$ ， 则称 X 与 Y 不相关。不相关是指 X 与 Y 之间没有线性关系， 但 X 与 Y 之间可能有其他的函数关系， 譬如平方关系、对数关

系等。

- 若 $Corr(X,Y)=1$，则称 X 与 Y 完全正相关；若 $\mathrm{Corr}(X,Y)=-1$，则称 X 与 Y 完全负相关。
- 若 $0<|Corr(X,Y)|<1$，则称 X 与 Y 有“一定程度”的线性关系。$|Corr(X,Y)|$ 越接近于 1，则线性相关程度越高；$|Corr(X,Y)|$ 越接近于 0，则线性相关程度越低。而协方差看不出这一点，若协方差很小，而其两个标准差 σ_X 和 σ_Y 也很小，则其比值就不一定很小。

4. 大数定律与中心极限定理

随机变量序列的收敛性有很多种，其中常用的是依概率收敛和按分布收敛。大数定律涉及的是一种依概率收敛，中心极限定理涉及到按分布收敛。接下来我们来介绍一下。

4.1. 随机变量序列的两种收敛性

4.1.1. 依概率收敛

在第一部分我们使用频率确定概率时，我们提出“概率是频率的稳定值”，或“频率稳定于概率”，现在我们来解释“稳定”的含义。

设有一大批产品，其不合格品率为 p。现一个接一个地检查产品的合格性，记前 n 次检查发现 S_n 个不合格品, 而 $v_n=\dfrac{S_n}{n}$ 为不合格品出现的频率。当检查继续下去，我们就发现频率序列 $\{v_n\}$ 有如下两个现象：

（1）频率 v_n 对概率 p 的绝对偏差 $|v_n-p|$ 将随 n 增大而呈现逐渐减小的趋势，但无法说它收敛于零。

（2）由于频率的随机性，绝对偏差 $|v_n-p|$ 时大时小。虽然我们无法排除大偏差发生的可能性, 但随着n不断增大，大偏差发生的可能性会越来越小。这是一种新的极限概念。

用数学式子表示如下：对任意给定的$\varepsilon>0$，事件$\{|v_n-p|\geq\varepsilon\}$出现了就认为大偏差发生了，而大偏差发生的可能性越来越小，相当于$P(|v_n-p|\geq\varepsilon)\to 0(n\to\infty)$，这时就可以称序列$\{v_n\}$依概率收敛，这就是“频率稳定于概率”。

下面给出一般的随机变量序列$\{X_n\}$依概率收敛于一个随机变量X的定义：设$\{X_n\}$为随机变量序列，X为以随机变量，如果对任意的$\varepsilon > 0$，有$P(|X_n - X| \geq \varepsilon) \to 0 (n \to \infty)$。则称序列$\{X_n\}$依概率收敛于$X$，记作$X_n \xrightarrow{P} X$。其含义为：$X_n$对$X$的绝对偏差不小于任一给定量的可能性将随着$n$增大而减小。或者说，绝对偏差$|X_n - X|$小于任一给定量的可能性将随着$n$增大而越来越接近1，即$P(|X_n - X| < \varepsilon) \to 1 (n \to \infty)$。特别当$X$为退化分布时，即$P(X = c) = 1$，则称序列$\{X_n\}$依概率收敛于$c$，记作$X_n \xrightarrow{P} c$。

4.1.2. 按分布收敛

依概率收敛，描述的是当$n \to \infty$时，随机变量序列越来越接近（趋近于）某个确定的随机变量的概率接近于1。我们知道随机变量的分布函数全面描述了随机变量的统计规律，如何来定义$\{F_n(x)\}$的收敛性呢？可以猜想：对于所有的x，有$F_n(x) \to F(x) (n \to \infty)$呢？这其实就是按分布收敛，具体来说，对于随机变量$X, X_1, X_2, \cdots$的分布函数分别为$F(x), F_1(x), F_2(x), \cdots$。若对$F(x)$的任一连续点$x$，都有$\lim\limits_{n \to \infty} F_n(x) = F(x)$，则称$\{F_n(x)\}$弱收敛于$F(x)$，记作$F_n(x) \xrightarrow{W} F(x)$，也称相应的随机变量序列$\{X_n\}$按分布收敛于$X$，记作$X_n \xrightarrow{L} X$。

4.2. 大数定律

4.2.1. 伯努利大数定律

记S_n为n重伯努利试验中事件A出现的次数，称$\dfrac{S_n}{n}$为事件A出现的频率。如果记一次试验中A发生的概率为p，则S_n服从二项分布$b(n, p)$，因此频率$\dfrac{S_n}{n}$的数学期望和方差分别为

$$E(\frac{S_n}{n}) = p, \quad Var(\frac{S_n}{n}) = \frac{p(1-p)}{n}$$

,

对任意$\varepsilon > 0$，有$\lim\limits_{n \to \infty} P\left(|\dfrac{S_n}{n} - p| < \varepsilon\right) = 1$。

伯努利大数定律说明：随着n的增大，事件A发生的概率$\dfrac{S_n}{n}$与其概率p的偏差$|\dfrac{S_n}{n} - p|$大于预先给定的精度ε的可能性愈来愈小，这就是频率稳定于概率的含义。

下面我们来看一个例子，使用蒙特卡罗法计算定积分。设$0 \le f(x) \le 1$，求$f(x)$在区间$[0,1]$上的积分值$\int_0^1 f(x)dx$。

设二维随机变量(X,Y)服从正方形$\{0 \le x \le 1, 0 \le y \le 1\}$的均匀分布，则可知$X \sim U(0,1)$, $Y \sim U(0,1)$，且X与Y相互独立，记事件$A=\{Y \le f(X)\}$，则有

$$P(A) = P(Y \le f(X)) = \int_0^1 \int_0^{f(x)} dydx = \int_0^1 f(x)dx = J$$

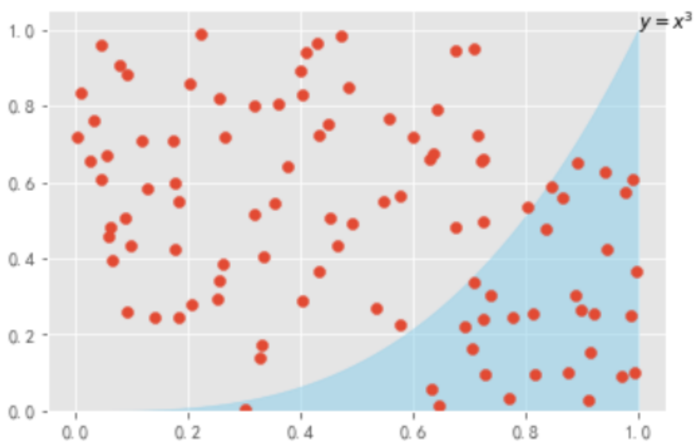
也就是定积分的值就是事件A的概率。由伯努利大数定律，我们可以使用重复试验中A出现的频率作为p的估计值。这种求解定积分的方法额称为随机投点法。

下面我们使用蒙特卡洛方法得到A出现的频率：

- 先产生 $(0,1)$ 上均匀分布的 $2n$ 个随机数，组成 n 对随机数 (x_i, y_i), $i=1,2,\cdots,n$， n 可以取很大的数
- 对 n 对数据 (x_i, y_i)，记录满足如下不等式 $y_i \le f(x_i)$ 的次数，这就是事件 A 发生的频数 S_n，由此可以计算事件 A 发生的频率 $\frac{S_n}{n}$，则积分 $\approx \frac{S_n}{n}$ 。

我们取$f(x)=x^3$，尝试使用蒙特卡洛投点法计算积分，代码如下：

```python
from scipy.stats import uniform
# 蒙特卡洛积分计算的原理:
x_arr = np.linspace(0,1,1000)
x_n = uniform.rvs(size = 100)  # 随机选择n个x随机数
y_n = uniform.rvs(size = 100)  # 随机选择n个y随机数
plt.stackplot(x_arr,x_arr ** 3,alpha=0.5,color="skyblue") #堆积面积图
plt.scatter(x_n,y_n)
plt.text(1.0,1.0,r'$y=x^3$')
plt.show()
```



```python
def montecarlo_method(n):
    x_n = uniform.rvs(size = n)  # 随机选择n个x随机数
    y_n = uniform.rvs(size = n)  # 随机选择n个y随机数
    fx = x_n ** 3
    sn = np.sum([y_n[i] <= fx[i] for i in range(n)])
    return sn / n

x = symbols("x")
print(f'y=x**3在[0,1]的定积分为: {integrate(x ** 3, (x,0, 1))}')
for n in [10, 100, 1000, 10000, 100000]:
    print(f'蒙特卡洛法模拟{n}次得到的积分近似值: {montecarlo_method(n)}')
```

```
y=x**3在[0,1]的定积分为: 1/4
蒙特卡洛法模拟10次得到的积分近似值: 0.2
蒙特卡洛法模拟100次得到的积分近似值: 0.23
蒙特卡洛法模拟1000次得到的积分近似值: 0.271
蒙特卡洛法模拟10000次得到的积分近似值: 0.2535
蒙特卡洛法模拟100000次得到的积分近似值: 0.25243
```

4.2.2. 辛钦大数定律

设 $\{X_n\}$ 为一独立同分布的随机变量序列，若 X_i 的数学期望存在，则 $\{X_n\}$ 服从大数定律，即对任意的 $\varepsilon > 0$，

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \varepsilon\right) = 1$$

成立。

对于独立同分布且具有相同均值 μ 的随机变量X，$X_1, X_2, \ldots\ldots X_n$，当 n 很大时，它们的算术平均数 $\frac{1}{n}\sum_{i=1}^{n} X_i$ 很接近于 μ。也就是说可以使用样本的均值去估计总体均值。

由辛钦大数定律我们可以得出，如果$\{X_n\}$为某一独立同分布的随机变量序列，且$E(|X_i|^k)$存在，其中k为正整数，则$\{X_n^k\}$服从大数定律，这就是说，我们其实可以将$\frac{1}{n}\sum_{i=1}^{n} X_i^k$作为$E(X_i^k)$的近似值。

因此，我们也可以用平均值法下的蒙特卡洛方法计算定积分$J = \int_0^1 f(x)dx$。设随机变量$X \sim U(0,1)$，则$Y = f(x)$的数学期望为$E(f(x)) = \int_0^1 f(x)dx$。由辛钦大数定律，可以用$f(X)$的观察值的平均取估计$f(X)$的数学期望的值。具体做法如下：

- 产生 n 个均匀分布的随机数 x_i，$i = 1, 2, \cdots, n$

- 对每个 x_i 计算 $f(x_i)$，最后得到 $J \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i)$

代码如下所示：

```
def montecarlo_method_mean(n):
    x_n = uniform.rvs(size=n)
    fx = x_n ** 3
    return np.mean(fx)

x = symbols("x")
print(f'y=x**3在[0,1]的定积分为: {integrate(x ** 3, (x,0, 1))}')
for n in [10, 100, 1000, 10000, 100000]:
    print(f'蒙特卡洛法(平均值法)模拟{n}次得到的积分近似值: {montecarlo_method_mean(n)}')
```

```
y=x**3在[0,1]的定积分为：1/4
蒙特卡洛法(平均值法)模拟10次得到的积分近似值：0.35391983862405274
蒙特卡洛法(平均值法)模拟100次得到的积分近似值：0.2708569337984175
蒙特卡洛法(平均值法)模拟1000次得到的积分近似值：0.2665572316943115
蒙特卡洛法(平均值法)模拟10000次得到的积分近似值：0.24909670562827538
蒙特卡洛法(平均值法)模拟100000次得到的积分近似值：0.25138494129256245
```

4.3. 中心极限定理

大数定律讨论的是在什么条件下，随机变量序列的算术平均依概率收敛到其均值的算术平均，现在我们讨论在什么条件下，独立随机变量和$Y_n=\sum_{i=1}^{n} X_i$的分布函数会收敛于正态分布。

我们先给出一个独立随机变量和的例子。操作者在机床上加工机械轴，由于加工时会受到一些随机因素的影响，因此会每个机械轴的直径产生误差，若将这个误差记为Y_n，那么随机变量Y_n可以看作很多微小的随机波动$X_1, X_2, \cdots, X_n$之和，即$Y_n = X_1 + X_2 + \cdots + X_n$。这里的$n$是很大的，当$n \to \infty$时，$Y_n$的分布是什么？

我们尝试用正态分布、均匀分布、指数分布、泊松分布、0-1分布来模拟

4.3.1. 正态分布

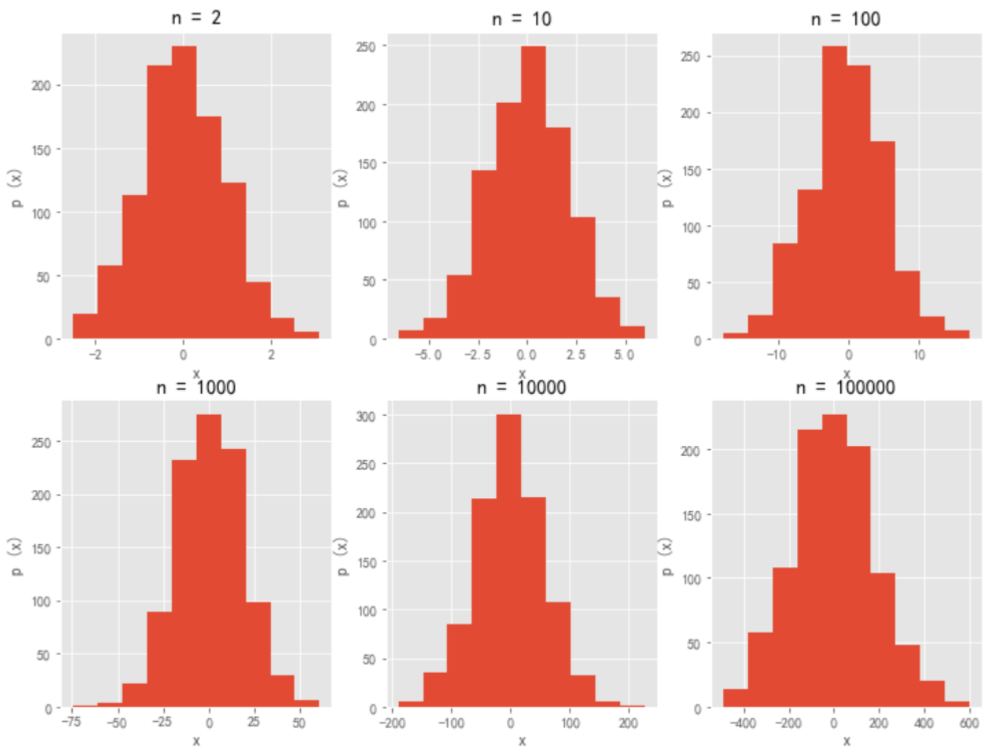
```
# 模拟n个正态分布的和的分布
from scipy.stats import norm

def plot_hist(arr, ax, n):
    ax.hist(arr)
    ax.set_title("n = "+str(n))
```

```python
    ax.set_xlabel("x")
    ax.set_ylabel("p (x)")

def norm_sum_n(n):
    num_samples = 1000
    arr = np.zeros(num_samples)
    for i in range(n):
        mu = 0
        sigma2 = np.random.rand()
        err_arr = norm.rvs(mu, sigma2, num_samples)
        arr += err_arr
    return arr


fig, axes = plt.subplots(2, 3, figsize=(12, 9))
for i, n in enumerate([2, 10, 100, 1000, 10000, 100000]):
    arr = norm_sum_n(n)
    plot_hist(arr, axes[i // 3][i % 3], n)
```



4.3.2. 均匀分布

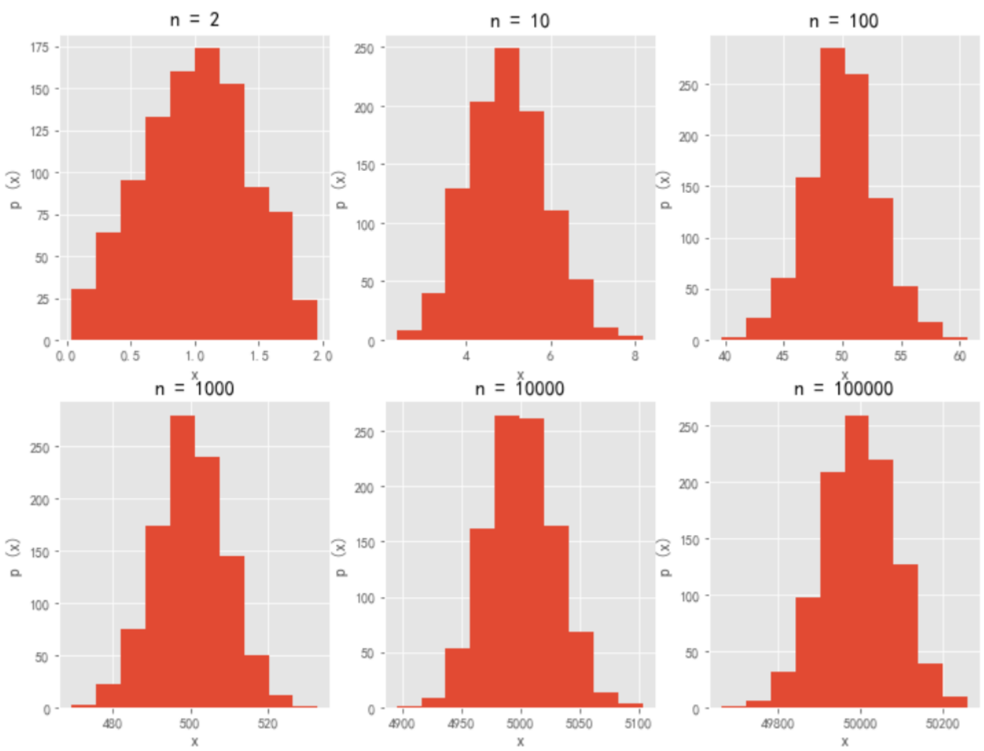
```python
# 模拟n个均匀分布的和的分布
from scipy.stats import uniform

def plot_hist(arr, ax, n):
    ax.hist(arr)
```

```python
    ax.set_title("n = "+str(n))
    ax.set_xlabel("x")
    ax.set_ylabel("p (x)")

def uniform_sum_n(n):
    num_samples = 1000
    arr = np.zeros(num_samples)
    for i in range(n):
        err_arr = uniform.rvs(size=num_samples)
        arr += err_arr
    return arr


fig, axes = plt.subplots(2, 3, figsize=(12, 9))
for i, n in enumerate([2, 10, 100, 1000, 10000, 100000]):
    arr = uniform_sum_n(n)
    plot_hist(arr, axes[i // 3][i % 3], n)
```



4.3.3. 指数分布

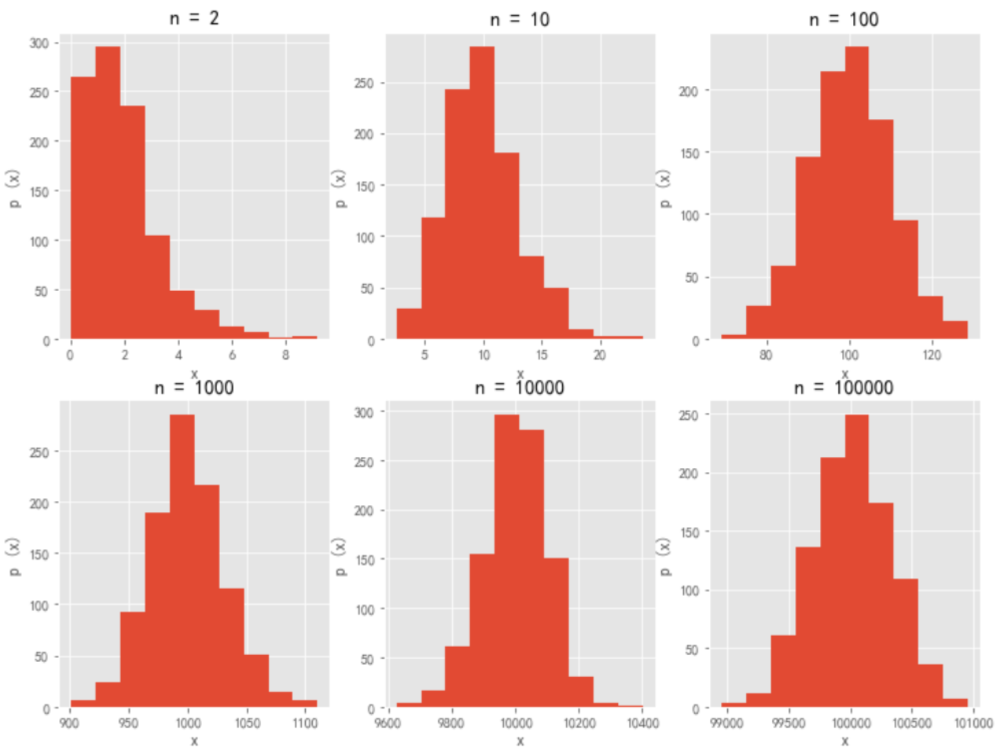
```python
# 模拟n个指数分布的和的分布
from scipy.stats import expon

def plot_hist(arr, ax, n):
    ax.hist(arr)
    ax.set_title("n = "+str(n))
    ax.set_xlabel("x")
```

```python
    ax.set_ylabel("p (x)")

def expon_sum_n(n):
    num_samples = 1000
    arr = np.zeros(num_samples)
    for i in range(n):
        err_arr = expon.rvs(size=num_samples)
        arr += err_arr
    return arr


fig, axes = plt.subplots(2, 3, figsize=(12, 9))
for i, n in enumerate([2, 10, 100, 1000, 10000, 100000]):
    arr = expon_sum_n(n)
    plot_hist(arr, axes[i // 3][i % 3], n)
```



4.3.4. 泊松分布

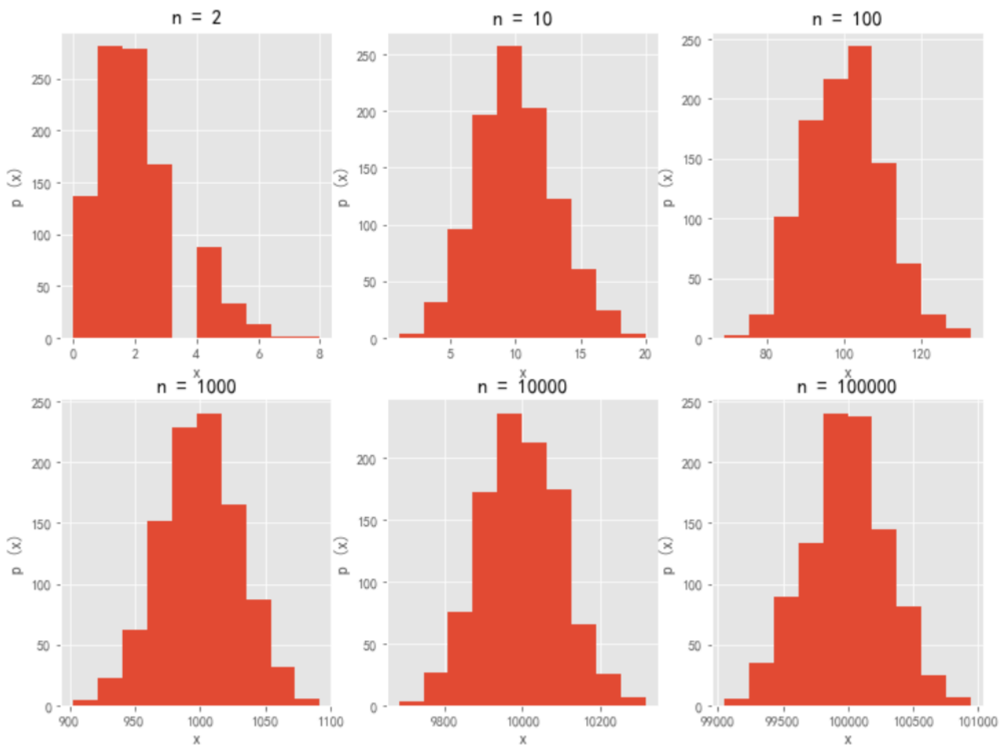
```python
# 模拟n个泊松分布的和的分布
from scipy.stats import poisson

def plot_hist(arr, ax, n):
    ax.hist(arr)
    ax.set_title("n = "+str(n))
    ax.set_xlabel("x")
    ax.set_ylabel("p (x)")

```

```python
def poisson_sum_n(n):
    num_samples = 1000
    arr = np.zeros(num_samples)
    for i in range(n):
        err_arr = poisson.rvs(mu=1.0, size=num_samples)
        arr += err_arr
    return arr


fig, axes = plt.subplots(2, 3, figsize=(12, 9))
for i, n in enumerate([2, 10, 100, 1000, 10000, 100000]):
    arr = poisson_sum_n(n)
    plot_hist(arr, axes[i // 3][i % 3], n)
```



4.3.5. 0-1分布

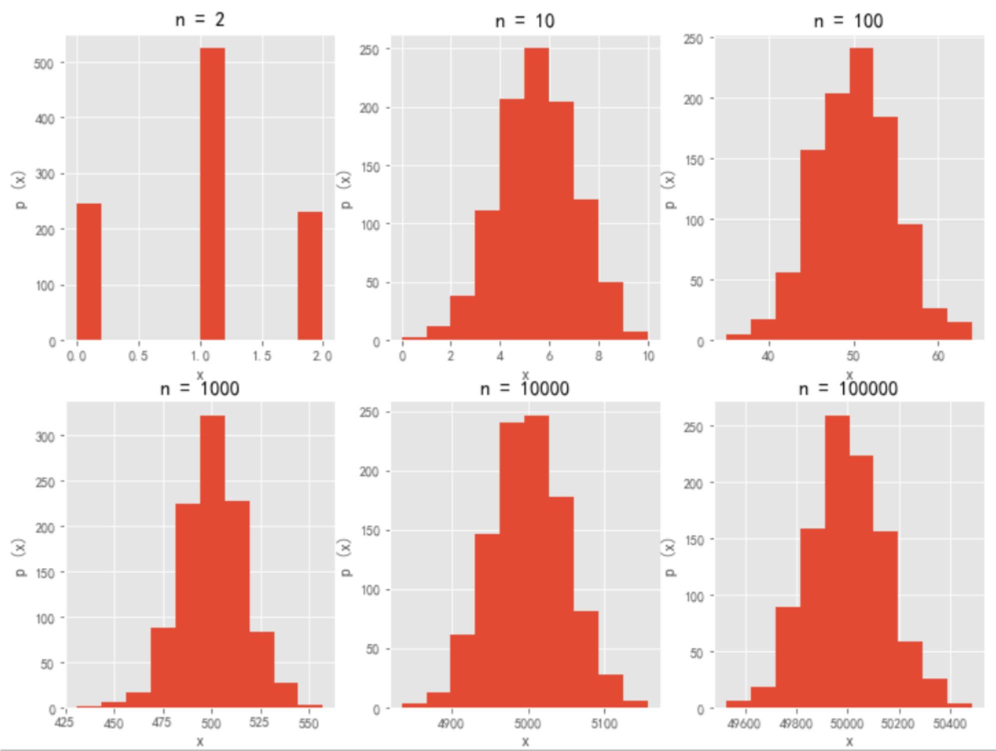
```python
# 模拟n个0-1分布的和的分布
from scipy.stats import bernoulli

def plot_hist(arr, ax, n):
    ax.hist(arr)
    ax.set_title("n = "+str(n))
    ax.set_xlabel("x")
    ax.set_ylabel("p (x)")

def bernoulli_sum_n(n):
    num_samples = 1000
```

```python
    arr = np.zeros(num_samples)
    for i in range(n):
        err_arr = bernoulli.rvs(p=0.5, size=num_samples)
        arr += err_arr
    return arr


fig, axes = plt.subplots(2, 3, figsize=(12, 9))
for i, n in enumerate([2, 10, 100, 1000, 10000, 100000]):
    arr = bernoulli_sum_n(n)
    plot_hist(arr, axes[i // 3][i % 3], n)
```



以上模拟说明：假设 $\{X_n\}$ 独立同分布、方差存在， 不管原来的分布是什么，只要 n 充分大，就可以用正态分布去逼近随机变量和的分布，这就是中心极限定理。

5. 数学建模综合案例分析：投资组合风险分析

GitModel公司是一家专业的投资银行，志在帮助客户更好地管理资产。客户手头上有一笔100万的资金，希望将这笔钱投入股票市场进行投资理财，投资人看中了两个股票A、B，股票分析师通过对股票A、B的历史数据分析发现：股票A的平均收益近似服从$N(0.1, 0.01)$，股票B的平均收益近似服从$N(0.3, 0.04)$。现在客户希望通过

分析得出投资股票A、B的最佳组合（在预期收益确定情况下最小风险时，需要投资A、B的份额）。

分析：首先，我们先来分析投资组合的收益应该如何计算：设A、B的投资收益率为随机变量X、Y，因此$X \sim N(0.1, 0.01)$，$Y \sim N(0.3, 0.04)$。设x_1为投资A的份额，$y_1 = 1 - x_1$为投资B的份额，因此投资组合的收益率为：$Z = x_1 * X + y_1 * Y$，投资组合的平均收益率为：$E(Z) = x_1 * E(X) + y_1 * E(Y)$。

接下来，我们来分析投资组合的风险应该如何计算：何为风险，最简单来说就是收益的不确定性，如果收益是确定且固定的，就无所谓的风险可言。根据对风险的直观描述，我们可以定义风险为收益率的方差，因此：股票A的风险为$\sigma_x^2 = 0.01$，股票B的风险为$\sigma_y^2 = 0.04$,而投资组合的风险为

$$
\begin{aligned}
Var(Z) &= Var(x_1 * X + y_1 * Y) \\
&= x_1^2 \operatorname{Var}(X) + y_1^2 \operatorname{Var}(Y) + 2x_1y_1 \operatorname{Cov}(X, Y)
\end{aligned}
$$

因此，最佳的投资组合应该是风险最小时的投资组合，即：

$$
\begin{aligned}
&min \quad Var(Z) \\
&= min \quad x_1^2 \operatorname{Var}(X) + y_1^2 \operatorname{Var}(Y) + 2x_1y_1 \operatorname{Cov}(X, Y) \\
&= \frac{d(Var(Z))}{d(x_1)} = 0
\end{aligned}
$$

我们尝试用代码来求解，这里我们取相关系数$\rho = 0.4$：

```python
from sympy import *
x = symbols('x')
y = symbols('y')
y = 1-x
## 请根据var_Z的定义写出相应的公式代码
var_z = x * x * 0.01 + y * y * 0.04 + 2 * x * y * 0.4 * 0.1 * 0.2
## 一阶导数=0
derivation_var_z = diff(var_z, x)
stagnation = solve(derivation_var_z, x)
print(stagnation)
```

```
[0.941176470588235]
```

因此，根据风险最小化原则，应该投资A股票$10w \times 0.94$元，而投资B股票$10w \times 0.06$元。

6. 小结

这次的打卡内容相当的充实，也花费了不少时间。虽然已经对教材和教程内容做了一定的精简，但是篇幅还是相当大，当然本篇文章也只是给帮助大家回顾概率论所学知识，欢迎小伙伴一起交流~