# 3.5-A 推理学习：概率论基础

概率论和图论是概率图的基础，图论的基础可以查看*复杂网络*部分，本章主要汇集概率论的基础知识点，并利用Python相关库介入计算。以《概率论与数理统计》[1,2]为主要参考书目；以[datascience](https://github.com/data-8/datascience)①、[SymPy](https://docs.sympy.org/latest/index.html)②为主要计算工具；结合加州大学伯克利分校（University of California（UC）, Berkeley）[数据科学的概率课程（Probability for Data Science）在线教材](http://prob140.org/textbook/content/README.html)[3]整理书写。

## 3.5.1 随机事件与概率

* 随机试验

**概率论**是一门研究**随机现象**（事先无法预知会出现哪个结果）统计规律性的数学学科。随机现象在一次试验中呈现不确定性的结果，而在大量重复试验中结果将呈现某种规律性，称为**统计规律性**。为研究随机现象的统计规律性，对客观事物进行观察的过程为**随机试验（试验）**。随机试验的特点有：

（1）在相同条件下试验可以重复进行；

（2）每次试验的结果不止一种，但是试验之前必须明确所有可能的结果；

（3）每次试验将会出现什么样的结果事先无法预知。

* 样本空间

随机试验一切可能结果组成的集合称为**样本空间**，记为，其中表示试验的每一个可能结果，又称为**样本点**，即样本空间为全体样本点的集合。样本空间中的元素可以是数（例如，投掷一枚均匀骰子的样本空间为），也可以不是数（例如，抛掷一枚硬币的样本空间为*Ω ={*H, T}*Ω={*c*:*c*⩾0}）。*

* 随机事件

通过随机试验研究随机现象时，每次试验都只能出现中的某一个结果（样本点）。各个可能结果是否在一次试验中出现是随机的。在随机试验中，常关注某些结果是否出现，例如空气质量指数（Air quality index，AQI）是否超出了评定空气质量为优的阈值，掷出点数是否为奇数（可描述为，为的一个子集）等。这些在一次试验中可能出现和可能不出现的一类结果称为**随机事件（事件）**，常用大写字母等表示。从集合的角度，随机事件为样本空间部分样本点组成的集合。在事件的定义中，有如下概念：

（1）任意随机事件是样本空间的一个子集；

（2）当试验结果属于该子集时，就说事件发生了；否则，不属于该子集，事件没有发生；

（3）仅含一个样本点的随机事件称为**基本事件**；

（4）样本空间也是自己的一个子集，也称为一个事件。因为包含所有可能的试验结果，所以在每次试验中一定发生，又称为**必然事件**；

（5）空集也是样本空间的一个子集，所以也称为一个事件。因为中不包含任何元素，所以在每次试验中一定不发生，又称为**不可能事件**。

* 关系运算

因为随机事件为样本空间部分样本点组成的集合，因此同集合，给定一个随机试验，是它的样本空间，都为的子集，随机事件有如下关系和运算：

（1） 如果（或），则称事件包含在中（或称包含）。从概率论角度来说，事件发生必然导致事件发生；

（2）如果，同时成立，则称事件与相等，记为。从概率论角度来说，事件发生必然导致事件发生；同样，事件发生必然导致事件发生，即和为同一事件；

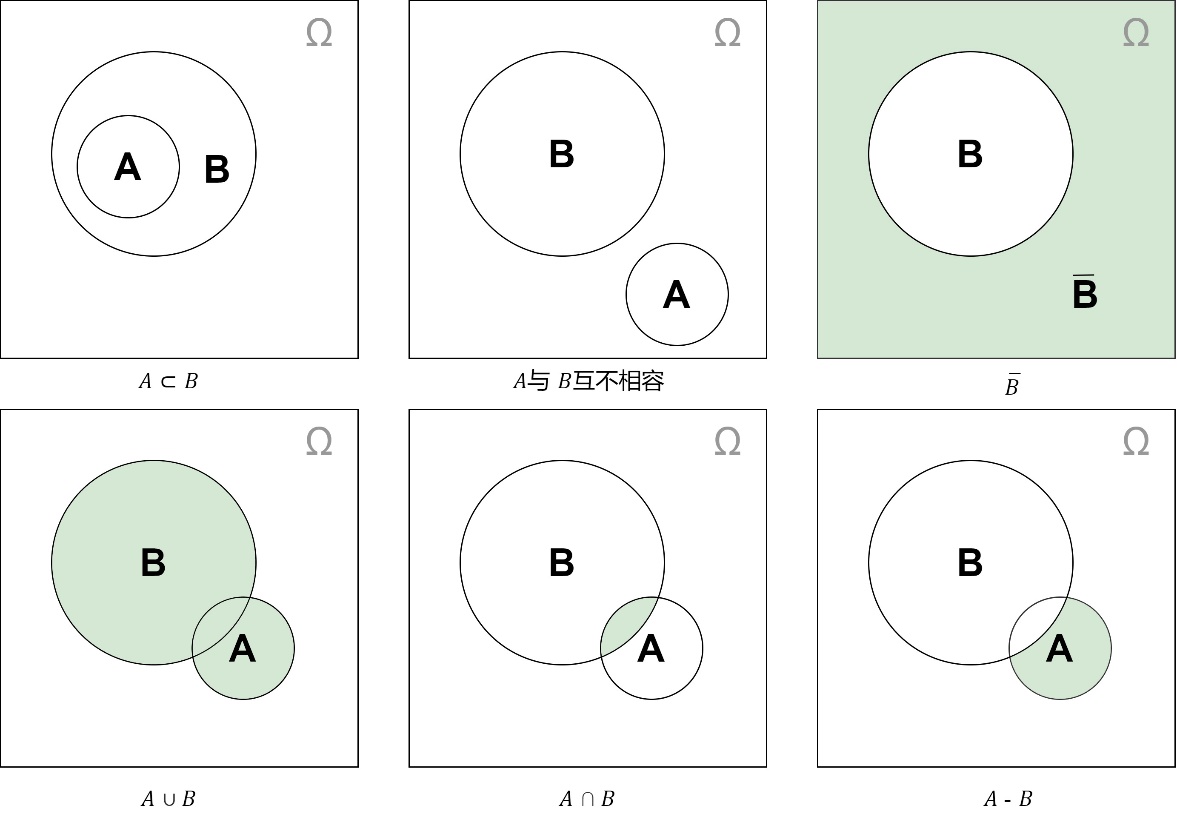
（3）如果和没有相同的样本点，则称事件和互不相容（或称为互斥），从概率论角度来说，事件和事件不可能同时发生；

（4）事件与的并，记为，表示由事件与中所有样本点组成的新事件，从概率论角度，事件与中至少有一个可能结果发生；

（5）事件与的交，记为（或），表示由事件和中公共的样本点组成的事件，从概率论角度，事件与同时发生；

（6）事件与的差，记为，表示由在事件中且不在事件中的样本点组成的新事件，从概率论角度，事件发生，且事件不发生；

（7）事件的对立事件（或称为逆事件、余事件），记为 ΩA A不发生。



从随机事件的关系和运算可以得出：

（1）对立事件一定是互不相容的事件，即，但互不相容事件不一定是对立事件；

（2）根据差事件和对立事件的定义，事件与的差还可以表示成；

（3）必然事件与不可能事件互为对立事件，即。

事件的运算性质，满足如下定律：

（1）交换律，A∪B=B∪A,A∩B=B∩A；

（2）结合律，(A∪B)∪C=A∪(B∪C),(AB)C=A(BC)；

（3）分配律，;

（4）对偶律，。

* 概率定义

在次试验中，如果事件出现了次，则称比值 nA AnAEΩAP(A)P(A)A的概率：

（1）**非负性公理**，对于任意事件，总有 ;

（2） **规范性公理**，；

（3）**可列可加性公理**，若A1, A2, , An。

由概率的3条公理，得到概率的一些重要基本性质如下：

（1）P(Ø)=0；

（2）有限可加性。设A1, A2, , An;

（3）对任意事件，有；

（4）若事件，则；

（5）减法公式。设为任意事件，则；

（6）加法公式。设为任意事件，则。

* 等可能概型

最简单的一种随机现象是样本空间中的每个基本事件发生的可能性相等，称之为**等可能概型**。例如抛掷一枚均匀的硬币或一颗均匀的骰子等。研究这一类随机现象的数学模型称之为**古典概型**。当样本空间是某个区域（可以是一维区间、二维平面或三维空间），例如搭乘地铁的等待时间、蒲丰投针问题等，称之为**几何概型**。

对于古典概型有，（1）随机试验的样本空间只有有限个样本点，可记作；（2）每个基本事件发生的可能性相等，即。若事件中含有个样本点，则事件的概率为： n\_AAnA={(1,1),(1,2), ,(1,6),(2,1), ,(6,6)}A={(1,1),(2,2), ,(6,6)}。

几何概型是古典概型的推广，保留每个样本点发生的等可能性，但去掉了中包含有限个样本点的限制，即允许试验可能结果有无穷不可列个。对几何概型有，（1）随机试验的样本空间是某个区域（可以是一维区间、二维平面或三维空间）；（2）每个样本点发生的可能性相等，则事件的概率为：m( ∙)

A，然后计算出相关图形的度量（一般为长度、面积或体积）。

* 条件概率

**条件概率**是指在某随机事件发生的条件下，另一随机事件发生的概率，记为，它与是不同的两类概率。例如，假设抛掷一枚均匀的骰子，已知掷出的点数是偶数，求点数超过3的概率。因为该试验的样本空间是Ω ={1,2,3,4,5 ,6}A={2,4,6}B={4,5,6}AB{4,6}AB 。

定义 1：设是随机试验，是样本空间，是随机试验上的两个随机事件且，称 ABP(B∣A)。

可以验证，条件概率同样满足概率公理化定义的三条基本性质，设，有（必须在同一条件下进行）：

（1）**非负性公理**，对于任意事件，总有；

（2）**规范性公理**，;

（3）**可列可加性公理**，若A1​,A2​,⋯。

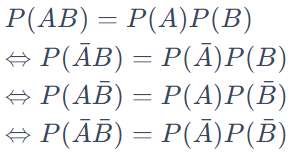
定理 1（概率的乘法公式）：设为随机试验上的两个事件，且，则有；同理，若，有。将其推广到多个事件的情况，设为任意的3个事件，且，有。更一般的表述为，设A1​,A2​,⋯,An​P(A1​A2​⋯An​)=P(A1​)P(A2​∣A1​)P(A3​∣A1​A2​)⋯P(An​∣A1​A2​⋯An−1​)。

·事件的相互独立性

一般来说，设为试验的两个事件，且，则事件的发生对事件发生的概率是有影响的，这时条件概率。但是，如果有，则可以推导出。即，不管事件发生还是不发生，都对事件发生的概率没有影响，可以理解为事件和事件之间没有“关系”，或者称事件和事件相互独立。

定义 2：设为试验的两个事件，如果满足等式，则称事件和事件**相互独立**，简称**独立**。

定理 2：若事件与事件相互独立，则下列各对事件也相互独立：与 B，即，



对于上4对事件，只要有一对是相互独立的，则其余3对也相互独立，可理解为：事件与相互独立，则的发生不会影响发生的概率；的发生也不会影响不发生的概率；的不发生也不会影响发生的概率；的不发生也不会影响不发生的概率。

将相互独立性推广到3各事件的情况，有，

定义 3：设是试验的三个事件，如果满足等式P(AB)=P(A)P(B)，P(AC)=P(A)P(C)，P(BC)=P(B)P(C)则称事件**两两相互独立**。

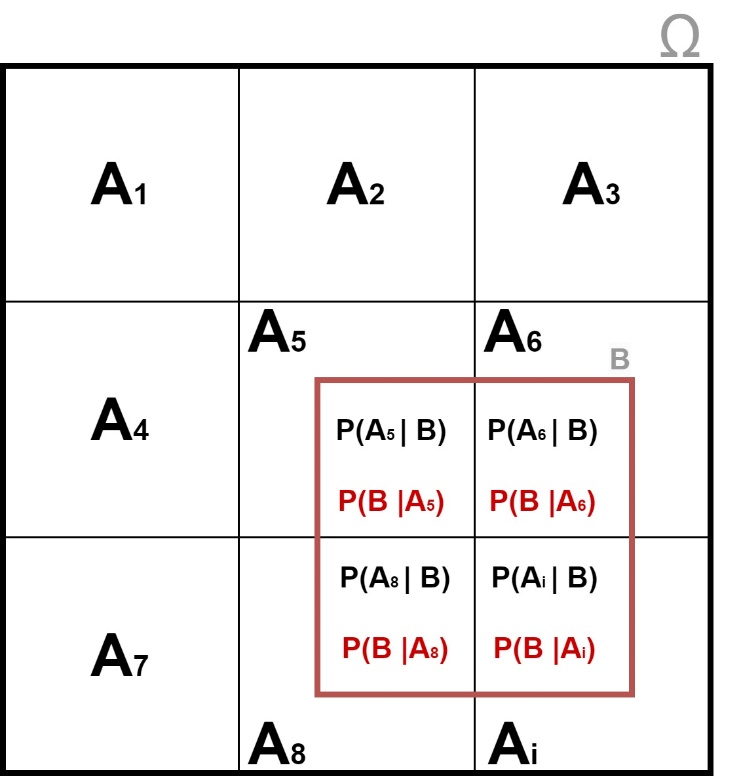
定义 4：设是试验的三个事件，如果满足等式  
P(AB)=P(A)P(B)，P(AC)=P(A)P(C)，P(BC)=P(B)P(C)，P(ABC)=P(A)P(B)P(C)则称事件**相互独立**。

* 全概率公式与贝叶斯公式

定义：设是随机试验，是相应的样本空间，A1​,A2​,⋯,An​A1​,A2​,⋯,An为样本空间的一个**完备事件组**，完备事件组完成了对样本空间的一个分割。

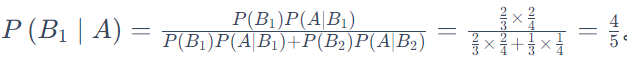
定理 1（**全概率公式**）：设  
A1​,A2​,⋯,An​ΩB。

定理 2（**贝叶斯公式**）：设A1​,A2​,⋯,An​BP(B)>0。



对全概率公式和贝叶斯公式可以通过上图理解，对于条件概率，即随机事件发生条件下，发生的概率的计算可以转化为条件概率的相关计算，即随机事件发生条件下，发生的概率。对全概率公式，通过已知每种“原因”发生的概率，即和已知，求结果发生的概率。这里和又称为**先验概率**；对于贝叶斯公式，则是从已知结果发生的条件下分析由各个可能“原因”引起的条件概率和，所以贝叶斯公式也可以看作“已知结果，分析原因”的问题，这里和又称为**后验概率**。

又例如，有两个口袋，甲袋中盛有2个白球，1个黑球；乙袋中盛有1个白球，2个黑球，由甲袋中任取一球放入乙袋，再从乙袋取出一球，问从乙袋中取出的球为白球的概率，则为根据以往的数据分析试验结果的概率，为先验概率，计算推导为：

。

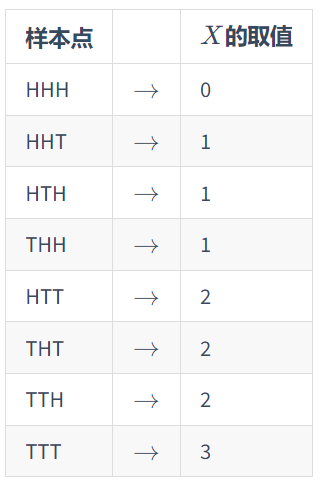
## 3.5.2 随机变量及其分布

* 随机变量的定义

定义 1：在随机试验中，是相应的样本空间，如果对中的每一个样本点，有唯一一个实数与它对应，那么就把这个定义域为的单值实值函数（Real-valued function）称为（一维）随机变量。例如，抛掷一枚均匀的硬币，观察其朝上的面，则样本空间（为正面，为反面），随机变量可设置为：

| 样本点 |  | 的取值 |
| --- | --- | --- |
| H | → | 1 |
| T | → | 0 |

如果抛掷三枚均匀的硬币，观察其朝上的面， 则样本空间Ω ={HHH,HHT,HTH,THH,HTT,THT,TTH,TTT}XX的取值与样本点之间的对应关系可设置为：



从上述例举可知，无论试验结果本身与数量是否有关，都可以把试验的每个结果与实数对应起来，即把试验结果数量化，由于这样的数量依赖于试验的结果，随试验结果的不同而变化，所以它的取值具有随机性，称这样的变量为随机变量，因此可以说，随机变量是试验结果的函数。

随机变量一般用大写字母等表示，随机变量的取值一般用小写字母等来表示。如果一个随机变量仅可能取有限或可列个值，则称其为离散型随机变量；如果一个随机变量的取值充满了数轴上的一个区间（或某几个区间的并），则称其为非离散型随机变量。连续型随机变量是非离散型随机变量中常见的一类随机变量。

随机变量定义的直观理解为，随机变量是样本点的函数，这个函数的自变量是样本点，可以是数，也可以不是数，定义域是样本空间；而因变量必须是实数，这个函数可以让不同的样本点对应不同的实数，也可以让多个样本点对应于同一个实数。

* 随机变量的分布函数

随机变量是样本点的一个实值函数，为了掌握的统计规律性，需要知道取值于某个区间的概率，由于{a<X≤b}={X≤b}−{X≤a},{X>c}=Ω−{X≤c}，因此对于任意实数，只需要知道的概率，用表示这个概率值。

定义 2：设是一个随机变量，对于任意实数，称函数F(x)=P(X≤x),−∞<x<+∞X−∞<a<b<+∞ XX(a,b]内的概率，即分布函数可以完整的描述一个随机变量的统计规律性。从定义可以得知：

（1）分布函数是定义在上，取值在上的一个函数；

（2）任意随机变量都有且仅有一个分布函数，有了分布函数，就可以计算与随机变量相关事件的概率问题。

* 离散型随机变量及其分布律

设是随机试验，是相应的样本空间，是上的随机变量，若的值域（记为）为有限集或可列集，此时称为（一维）**离散型随机变量**。

定义 3：若一维离散型随机变量的取值为  
x1​,x2​,…,xn​,…, P(X=xi​)=pi​,i=1,2,⋯ X的**分布律（或分布列、概率函数）**。

一维离散型随机变量的分布律可表示为：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 概率 |  |  |  |  |  |

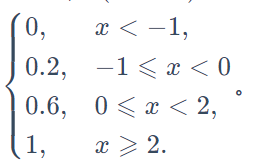
且满足（1）非负性pi​≥0,i=1,2,⋯ 。这两条性质也是判断某一数列是否能成为分布律的充要条件。

例如，设随机变量的分布律为：

|  | -1 | 0 | 2 |
| --- | --- | --- | --- |
| 概率 | 0.2 | 0.4 | 0.4 |

则可计算，；将分布律转化为分布函数有，x<−1,P(X⩽x)=0; −1⩽x<0,P(X⩽x)=P(X=−1)=0.2; 0⩽x<2,P(X⩽x)=P(X=−1)+P(X=0)=0.2+0.4=0.6;

x⩾2,P(X⩽x)=P(X=−1)+P(X=0)+P(X=2)=0.2+0.4+0.4=1. ，即F(x)=



已知一个离散型随机变量的分布律可以求得其分布函数，同样，已知一个离散型随机变量的分布函数可以求其分布律，有：P(X=−1)=P(X⩽−1)=F(−1)=0.2, P(X=0)=P(−1<X⩽0)=F(0)−F(−1)=0.6−0.2=0.4, P(X=2)=P(0<X⩽2)=F(2)−F(0)=1−0.6=0.4

分布函数和分布律对离散型随机变量的取值规律描述是等价的，分布律相对更直观方便。

* 连续型随机变量及其密度函数

连续型随机变量的取值充满了数轴上的一个或者某几个区间的并，在这个区间里有无穷不可列个实时，因此分布律不再适用描述连续型随机变量，而使用概率密度函数表示。

定义 4：设是随机试验，是相应的样本空间，是上的随机变量，是的分布函数，若存在非负函数使得，则称为（一维）连续型随机变量，称为的（概率）密度函数。其满足，（1）非负性；（2）规范性。

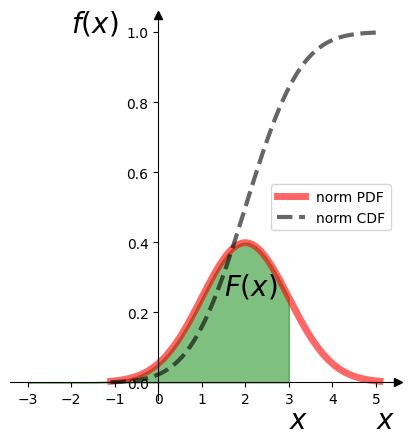
概率密度函数与分布函数之间的关系可用下图示意，恰好是在区间上的积分，为下图绿色填充部分的面积（或对应累积分布函数（CDF）的值）。连续型随机变量具有如下性质：

（1）分布函数是连续函数，在的连续点处，$ F’(x)=f(x) $；

（2）对任意一个常数c,−∞<c<+∞P(X=c)=0  
{a⩽X⩽b}X=aX=bP(a⩽X⩽b)=P(a<X⩽b)=P(a⩽X<b)=P(a<X<b)，而离散型随机变量计算的就是“点点概率”。

（3）对任意两个常数，有。

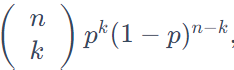
import numpy as np  
from scipy.stats import norm  
import matplotlib.pyplot as plt  
  
fig, ax=plt.subplots(1, 1,figsize=(5,5))  
  
loc,scale=2,1  
x=np.linspace(norm.ppf(0.001, loc, scale),norm.ppf(0.999, loc, scale), 100)  
ax.plot(x, norm.pdf(x, loc, scale),'r-', lw=5, alpha=0.6, label='norm PDF')  
ax.plot(x, norm.cdf(x, loc, scale),'k--', lw=3, alpha=0.6, label='norm CDF')  
  
px=np.arange(-3,3,0.001)  
ax.fill\_between(px,norm.pdf(px, loc, scale),alpha=0.5, color='g')  
  
ax.text(-2,1, r'$f(x)$', family="monospace",size=20)  
ax.text(1.5,0.25, r'$F(x)$', family="monospace",size=20)  
ax.text(5,-0.13, r'$x$', family="monospace",size=20)  
ax.text(3,-0.13, r'$x$', family="monospace",size=20)  
  
def plot\_style(ax):  
 # Move the left and bottom spines to x = 0 and y = 0, respectively.  
 ax.spines[["left", "bottom"]].set\_position(("data", 0))  
 # Hide the top and right spines.  
 ax.spines[["top", "right"]].set\_visible(False)  
 ax.plot(1, 0, ">k", transform=ax.get\_yaxis\_transform(), clip\_on=False)  
 ax.plot(0, 1, "^k", transform=ax.get\_xaxis\_transform(), clip\_on=False)  
  
plot\_style(ax)  
plt.legend(loc='center right')  
plt.show()



* 常用离散型随机变量

1）二项分布

若随机试验只有2个可能结果，和，则称为**伯努利试验（Bernoulli trial）**。设事件在一次试验中发生的概率，则。将独立的重复进行次，则称**重伯努利试验**。

记随机变量表示在重伯努利试验中事件发生的次数，可知的所有可能取值为。对每一，事件即为“次试验中事件恰好发生次”，由于各试验是相互独立的，根据事件的独立性，的分布律为： P(X=k)= , 0<p<1, k=0,1, , n Xn,p X X∼B(n,p)

nk(k≤n)nk

n=3k=2。

根据二项式展开公式：，可以推导出；且，，满足离散型随机变量分布律的规范性和非负性。

当时，X*∼*B*(1,*p*)*P(X=k)=(

) pk(1-p){1-k}=pk(1-p){1-k}, 0<p<1, k=0,1，相应的分布律为：

|  | 0 | 1 |
| --- | --- | --- |
| 概率 |  |  |

表示随机变量只取2个值，分别为 0 和 1，故又可称随机变量服从参数的分布（或两点分布）。

由服从参数为二项分布的随机变量的分布律计算公式，定义P\_Xk\_binomialDistributio()函数计算。例如某人向同一目标重复独立射击5次，每次命中目标的概率为0.8，求此人命中3次的概率为PX\_a=0.2048；此人至少命中2次的概率为PX\_b=0.99328。

# IPython extension to reload modules before executing user code.  
%load\_ext autoreload   
# Reload all modules (except those excluded by %aimport) every time before executing the Python code typed.  
%autoreload 2  
import usda.pgm as usda\_pgm  
import matplotlib.pyplot as plt  
import matplotlib  
import numpy as np  
from scipy.stats import poisson  
from scipy.stats import expon  
import math

The autoreload extension is already loaded. To reload it, use:  
 %reload\_ext autoreload

PX\_a=usda\_pgm.P\_Xk\_binomialDistribution(n=5,k=3,p=0.8)  
print(PX\_a)

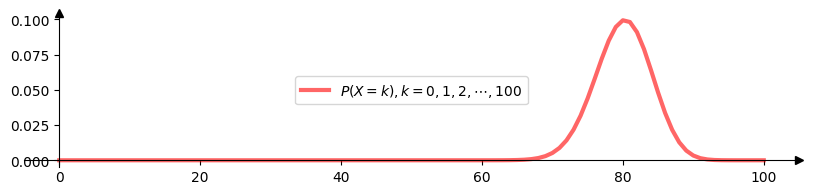
0.20479999999999993

PX\_b=1-usda\_pgm.P\_Xk\_binomialDistribution(n=5,k=0,p=0.8)-usda\_pgm.P\_Xk\_binomialDistribution(n=5,k=1,p=0.8)  
print(PX\_b)

0.99328

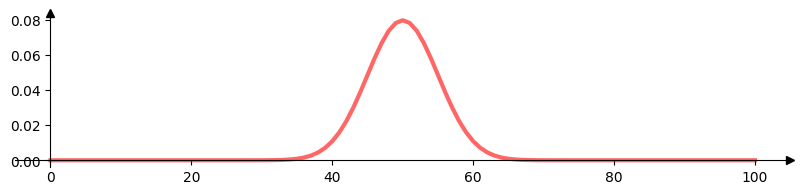
如果向同一目标重复独立射击100次，计算，可以发现二项分布近似于正态分布。

n=100  
PX\_100=[usda\_pgm.P\_Xk\_binomialDistribution(n=n,k=i,p=0.8) for i in range(n+1)]  
pts=[i for i in range(n+1)]  
  
fig, ax=plt.subplots(1, 1,figsize=(10,2))  
ax.plot(pts, PX\_100,'r-', lw=3, alpha=0.6, label='$P(X=k), k=0,1,2,\cdots, 100$')   
  
plot\_style(ax)  
plt.legend(loc=10)  
plt.show()



如果调整命中的概率为，则可以发现最大概率趋于约中间的位置。

PX\_100=[usda\_pgm.P\_Xk\_binomialDistribution(n=n,k=i,p=0.5) for i in range(n+1)]  
  
fig, ax=plt.subplots(1, 1,figsize=(10,2))  
ax.plot(pts, PX\_100,'r-', lw=3, alpha=0.6, label='$P(X=k), k=0,1,2,\cdots, 100$')  
  
plot\_style(ax)  
plt.show()



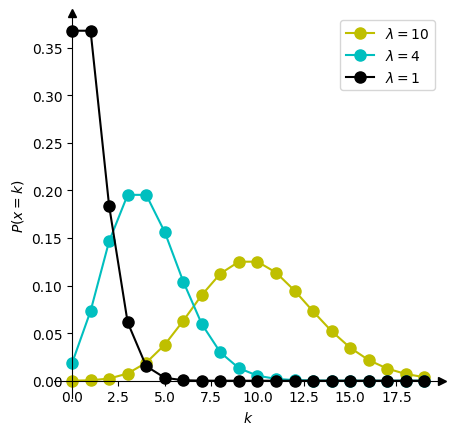
2） 泊松分布

设随机变量的取值为0,1,2, …,n,… , λ>0, k=0,1,2,… ,n, …XX∼P(λ)P{X=k} , k=0,1,2, …**。

泊松分布适合于描述单位时间（或空间）内随机事件发生的次数，如，某一服务设施在一定时间内到达的人数，电话交换机接到呼叫的次数，汽车站台的后客人数，机器出现的故障数，显微镜下单位分区内的细菌分布数等。

下述应用SciPy库提供的poisson方法绘制泊松分布，横轴为随机事件出现的次数（整数值），是预期的发生率，纵轴是给定时出现次的概率。

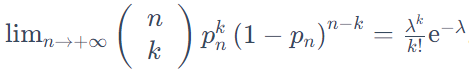
fig, ax=plt.subplots(1, 1,figsize=(5,5))  
x=np.arange(0,20) # x=np.arange(poisson.ppf(0.001, 10),poisson.ppf(0.999, 10))  
ax.plot(x, poisson.pmf(x, 10), 'yo-', ms=8, label='$\lambda=10$')  
ax.plot(x, poisson.pmf(x, 4), 'co-', ms=8, label='$\lambda=4$')  
ax.plot(x, poisson.pmf(x, 1), 'ko-', ms=8, label='$\lambda=1$')  
  
plot\_style(ax)  
plt.legend(loc='upper right')  
plt.xlabel("$k$")  
plt.ylabel("$P(x=k)$")  
plt.show()



例如，已知一购物网站每周销售某款商品的数量服从参数为6的泊松分布，问周初至少预备多少货源才能保证该周不脱销的概率不小于0.9（假设上周没有库存，且本周不再进货）。由poisson.ppf（Percent point function (inverse of cdf — percentiles)）方法可以直接计算（货源量），得。并用poisson.cdf（Cumulative distribution function，累积分布函数）验证时的概率，为0.916，不小于题目给定的0.9。

k=poisson.ppf(q=0.9, mu=6)  
print(k)  
p=poisson.cdf(k, 6)  
print(p)

9.0  
0.9160759830051242

泊松分布可以作为二项分布的一种近似，有**泊松定理**，在重伯努利试验中，记事件在一次试验中发生的概率为，如果当n→+∞ 。

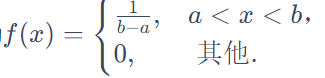
例如，设某保险公司的某类保险有1000人投保，每个投保人在1年内死亡的概率为0.005，且每个人在1年内是否死亡是相互独立的，试求在未来1年中者1000个投保人中死亡人数不超过10人的概率。因为较大，而较小，因此下述分别应用二项分布和泊松分布计算，结果近似相等。

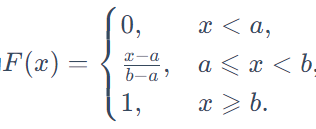
n=1000  
k=10  
p=0.005  
PX\_1=1-usda\_pgm.P\_Xk\_binomialDistribution(n=n,k=k,p=p)  
  
lamb=n\*p  
PX\_2=poisson.cdf(k, mu=lamb)  
print(f'PX\_1={PX\_1},\nPX\_2={PX\_2}')

PX\_1=0.9820037708176401,  
PX\_2=0.9863047314016171

* 常用连续型随机变量

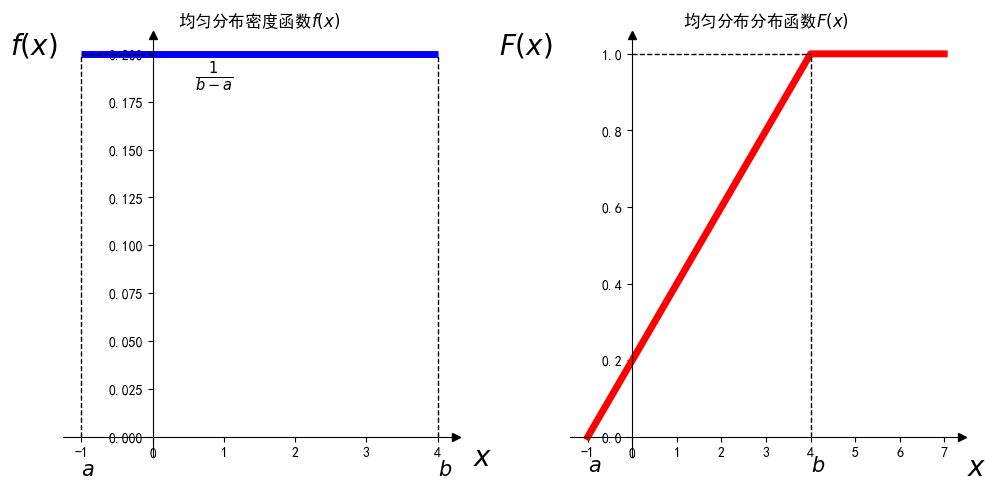
1）均匀分布

设为随机变量，对任意2个实数，概率密度函数为X(a,b)X*∼*U*(*a*,*b*)*。

若，则相应的分布函数（下图右）为

X(a,b)dc无关。

plt.rcParams['font.sans-serif'] = ['SimHei']  
matplotlib.rcParams['axes.unicode\_minus'] = False  
  
fig, axs=plt.subplots(1, 2,figsize=(10,5))  
  
a,b=-1,4  
axs[0].hlines(y =1/(b-a), xmin = a, xmax = b,linewidth=5, color='b')  
axs[0].vlines(x=[a,b],ymin=0,ymax =1/(b-a), linewidth=1, color='k',linestyles='dashed')  
axs[0].hlines(y =1/(b-a), xmin = 0, xmax = a,linewidth=1, color='k',linestyles='dashed')  
axs[0].text(-2,0.2, r'$f(x)$', family="monospace",size=20)  
axs[0].text(4.5,-0.015,'$x$', family="monospace",size=20)  
axs[0].text(0.6,1/(b-a)-0.015,'$\\frac{1}{b-a}$', family="monospace",size=15)  
axs[0].text(a,-0.02,'$a$', family="monospace",size=15)  
axs[0].text(b,-0.02,'$b$', family="monospace",size=15)  
axs[0].set\_title("均匀分布密度函数$f(x)$")  
  
axs[1].plot([a,b,b+3],[0,1,1],linewidth=5, color='r')  
axs[1].vlines(x=b,ymin=0,ymax =1, linewidth=1, color='k',linestyles='dashed')  
axs[1].hlines(y =1, xmin = 0, xmax = b,linewidth=1, color='k',linestyles='dashed')  
axs[1].text(-3,1, r'$F(x)$', family="monospace",size=20)  
axs[1].text(7.5,-0.1,'$x$', family="monospace",size=20)  
axs[1].text(a,-0.09,'$a$', family="monospace",size=15)  
axs[1].text(b,-0.09,'$b$', family="monospace",size=15)  
axs[1].set\_title("均匀分布分布函数$F(x)$")  
  
plot\_style(axs[0])  
plot\_style(axs[1])  
fig.tight\_layout()  
plt.show()



例如，设随机变量，求事件的概率。因为的概率密度函数为，所以。如果事件表示对作3次相互独立重复观测中事件出现的次数，则，所以。

2）指数分布

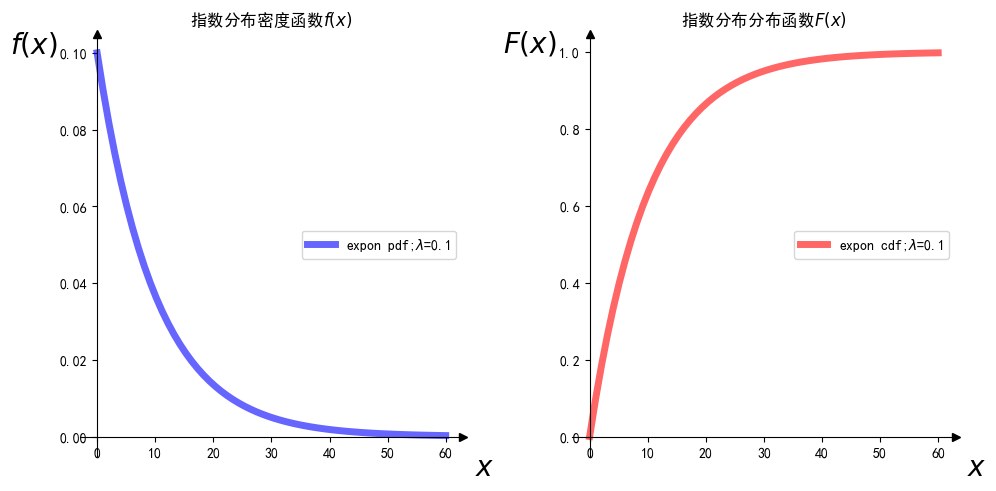
设为随机变量，概率密度函数（下图左）为，则称随机变量服从参数为的**指数分布**，记为。

若，则相应的分布函数（下图右）为$F(x)=

X∼E(λ), 0<a<b。

服从指数分布的随机变量只能取非负实数，常用于各种“寿命”分布，例如电子元件的寿命，某个特定事件发生所需的等待时间等往往服从指数分布。

def expon\_lamb\_pdf(x,lamb):  
 return lamb\*math.e\*\*(-lamb\*x)  
  
def expon\_lamb\_cdf(x,lamb):  
 return 1-math.e\*\*(-lamb\*x)  
  
plt.rcParams['font.sans-serif'] = ['SimHei']  
matplotlib.rcParams['axes.unicode\_minus'] = False  
  
fig, axs=plt.subplots(1, 2,figsize=(10,5))  
x=np.linspace(0,60, 60)  
lamb=0.1  
  
axs[0].plot(x, expon\_lamb\_pdf(x,lamb),'b-', lw=5, alpha=0.6, label=f'expon pdf;$\\lambda$={lamb}')  
axs[1].plot(x, expon\_lamb\_cdf(x,lamb),'r-', lw=5, alpha=0.6, label=f'expon cdf;$\\lambda$={lamb}')  
  
for ax in axs: ax.legend(loc='center right')  
axs[0].set\_title("指数分布密度函数$f(x)$")  
axs[1].set\_title("指数分布分布函数$F(x)$")  
axs[0].text(-15,0.1, r'$f(x)$', family="monospace",size=20)  
axs[0].text(65,-0.01,'$x$', family="monospace",size=20)  
axs[1].text(-15,1, r'$F(x)$', family="monospace",size=20)  
axs[1].text(65,-0.1,'$x$', family="monospace",size=20)  
  
plot\_style(axs[0])  
plot\_style(axs[1])  
fig.tight\_layout()  
plt.show()



3） 正态分布

设连续型随机变量的概率密度为，式中，μ,σ(σ>0)Xμ,σ  
X∼N(μ,σ2)。服从正态分布的随机变量称为**正态随机变量**。*连续型随机变量及其密度函数*部分给出的图即为正态分布的概率密度函数和分布函数。

* 离散型随机变量函数的分布

设为离散型随机变量的函数，则对的任一可能取值，有，。一般的，若是离散型随机变量，的分布律为：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |

则的分布律为：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |

* 连续型随机变量函数的分布

设是连续型随机变量，是连续函数，一般来说，Y=g(X)也是连续型随机变量，令分布为的概率密度函数，分别为的分布函数，则。利用的分布函数求出的概率即得的分布函数；若求的概率密度函数，只需对求导。

## 3.5.3 多维随机变量及其分布

* 多维随机变量

定义 1：设有随机试验，其样本空间为，若对中的每一个样本点都有一对有序实数与其对应，则称为**二维随机变量或二维随机向量**，称的取值范围为它的值域，记为。可以说二维随机变量是一个特殊的二元函数，其定义域为样本空间，值域Ω(X,Y)​⊂R²。

例：现有一枚骰子相互独立地上抛两次的随机试验，观察两次出现的点数，样本空间为如下计算结果Omega，总共36个样本点（两两组合的结果）。

import itertools  
from IPython.display import display, Latex

Omega=list(itertools.product(range(1,7),range(1,7)))  
print(len(Omega))  
display(Latex(f'$\Omega$={Omega}'))

36

=[(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)]

设随机变量表示“第1次出现的点数”，随机变量表示“两次出现点数的最小值”，并设为中的第个样本点，则对应样本点定义域转化为随机变量对应的值域，计算结果为XY，如下打印了一一对应结果。

XY=[(i[0],min(i)) for i in Omega]  
i=1  
for xy,omega in zip(XY,Omega):  
 display(Latex(f'$\omega\_{{{i}}}$={omega},$X(\omega\_{{{i}}}),Y(\omega\_{{{i}}})$={xy};'))   
 i+=1

=(1, 1),=(1, 1);

=(1, 2),=(1, 1);

=(1, 3),=(1, 1);

=(1, 4),=(1, 1);

=(1, 5),=(1, 1);

=(1, 6),=(1, 1);

=(2, 1),=(2, 1);

=(2, 2),=(2, 2);

=(2, 3),=(2, 2);

=(2, 4),=(2, 2);

=(2, 5),=(2, 2);

=(2, 6),=(2, 2);

=(3, 1),=(3, 1);

=(3, 2),=(3, 2);

=(3, 3),=(3, 3);

=(3, 4),=(3, 3);

=(3, 5),=(3, 3);

=(3, 6),=(3, 3);

=(4, 1),=(4, 1);

=(4, 2),=(4, 2);

=(4, 3),=(4, 3);

=(4, 4),=(4, 4);

=(4, 5),=(4, 4);

=(4, 6),=(4, 4);

=(5, 1),=(5, 1);

=(5, 2),=(5, 2);

=(5, 3),=(5, 3);

=(5, 4),=(5, 4);

=(5, 5),=(5, 5);

=(5, 6),=(5, 5);

=(6, 1),=(6, 1);

=(6, 2),=(6, 2);

=(6, 3),=(6, 3);

=(6, 4),=(6, 4);

=(6, 5),=(6, 5);

=(6, 6),=(6, 6);

上述对应样本点随机变量的值，存在重复值，将其转化为集合，仅保留唯一值，即为的值域，如下Omega\_XY，总计有21个值。

Omega\_XY=set(XY)  
print(len(Omega\_XY))  
display(Latex(f'$\Omega\_{{X,Y}}$={Omega\_XY}'))

21

={(4, 3), (3, 1), (5, 4), (5, 1), (2, 2), (6, 2), (6, 5), (4, 2), (3, 3), (5, 3), (2, 1), (6, 1), (6, 4), (3, 2), (4, 1), (5, 2), (4, 4), (5, 5), (1, 1), (6, 6), (6, 3)}

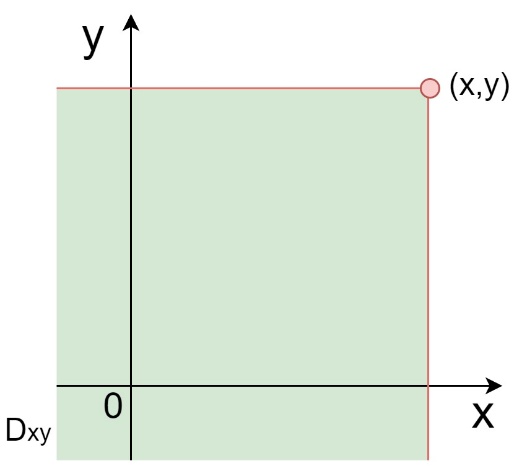
将二维随机变量的定义推广至维，

定义 2：设有随机试验，其样本空间为，若对中的每一个样本点都有一组有序实数列与其对应，则称为**维随机变量或维随机向量**，称的取值范围为它的**值域**，记为ΩX1​,X2​,⋯,Xn​​。

* 联合分布函数

二维随机变量，的分布不仅要包含每个随机变量各自的信息，还要包含两者之间相互关系的信息，因此称它们的分布为联合分布。

定义 3：设为二维随机变量，对任意的，称为随机变量的**（联合）分布函数**。式中，中的逗号表示对事件和事件取积事件，，式中区域如下图。联合分布函数描述了二维随机变量的统计规律，如果看作平面上随机点的坐标，那么其分布函数在处的函数值就表示随机变量落在以点为顶点且位于该点左下方的无穷矩形区域内的概率。



对于维随机变量的联合分布概率的定义。

定义 4：设为维随机变量，对任意的，称为**随机变量的（联合）分布函数**。

联合分布函数具有如下性质，

定理 1（**联合分布函数的性质**）：设是二维随机变量的联合分布函数，则，

（1）；

（2）当固定值时，是变量的非减函数，当固定值时，是变量的非减函数；

（3）；

（4）当固定值时，是变量的右连续函数，当固定值时，是变量的右连续函数；

（5）。

除了性质（5）,其它4条性质都可以推广至高维随机变量的联合分布函数。

* 二维离散型随机变量及其联合分布律

定义 5：如何二维随机变量仅可能取有限个或可列无限个值，则称为**二维离散型随机变量**。

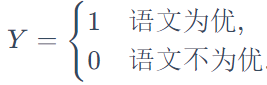
二维离散型随机变量的分布可用联合分布律表示。

定义 6：称为二维随机变量的**联合分布律**，式中，。

二维离散型随机变量的联合分布律可用表格发、公式法或图像法表示，其中最为简洁和常用的是表格法（Joint Distribution Table，联合分布表），如：

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

例如，分析一个年级的成绩分布，定义随机变量为，

(X,Y)P(Y=1)=0.1, P(X=1)=0.2,P(X=1,Y=1)=0.08P(X=0,Y=1)=P(Y=1)-P(X=1,Y=1)=0.02P(X=1,Y=0)=P(X=1)-P(X=1,Y=1)=0.12∑i​∑j​pij​=1 P(X=0,Y=0)=1-P(X=1,Y=1)-P(X=0,Y=1)-P(X=1,Y=0)=1-0.08-0.02-0.12=0.78。

使用[datascience](https://github.com/data-8/datascience)①库的Table()方法，根据计算结果建立联合分布表joint\_dist，计算结果如下。

from datascience import \*  
from prob140 import \*  
%matplotlib inline  
import matplotlib.pyplot as plt  
import numpy as np  
plt.style.use('fivethirtyeight')

k=np.arange(2)  
  
def joint\_probability(x,y):  
 if x==0 and y==0: return 0.78  
 elif x==0 and y==1: return 0.02  
 elif x==1 and y==0: return 0.12  
 elif x==1 and y==1: return 0.08  
  
joint\_dist=Table().values('X',k,'Y',k).probability\_function(joint\_probability)  
joint\_dist

|  | **X=0** | **X=1** |
| --- | --- | --- |
| **Y=1** | 0.02 | 0.08 |
| **Y=0** | 0.78 | 0.12 |

计算概率。使用定义的joint\_dist联合分布表，定义随机变量的关系函数indicator\_equal计算结果同。

def indicator\_equal(i, j):  
 return i <= j   
  
joint\_dist.event(indicator\_equal, 'X', 'Y')

P(Event) = 0.88

|  | **X=0** | **X=1** |
| --- | --- | --- |
| **Y=1** | 0.02 | 0.08 |
| **Y=0** | 0.78 |  |

的值域为，由得联合分布函数为：。

对*多维随机变量*部分将一枚骰子相互独立地上抛两次的随机试验，建立联合分布表（分布律）如下。

import collections  
import numpy as np  
  
k=np.arange(1,7)  
  
def joint\_probability(i,j):  
 n=len(XY)  
 frequency=dict(collections.Counter(XY))  
 if (i,j) in frequency.keys():  
 return frequency[(i,j)]/n  
 else:  
 return 0  
  
joint\_dist=Table().values('X',k,'Y',k).probability\_function(joint\_probability)  
joint\_dist

|  | **X=1** | **X=2** | **X=3** | **X=4** | **X=5** | **X=6** |
| --- | --- | --- | --- | --- | --- | --- |
| **Y=6** | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.027778 |
| **Y=5** | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.055556 | 0.027778 |
| **Y=4** | 0.000000 | 0.000000 | 0.000000 | 0.083333 | 0.027778 | 0.027778 |
| **Y=3** | 0.000000 | 0.000000 | 0.111111 | 0.027778 | 0.027778 | 0.027778 |
| **Y=2** | 0.000000 | 0.138889 | 0.027778 | 0.027778 | 0.027778 | 0.027778 |
| **Y=1** | 0.166667 | 0.027778 | 0.027778 | 0.027778 | 0.027778 | 0.027778 |

计算，代码方法如下。

joint\_dist.event(lambda i,j:i==j, 'X', 'Y')

P(Event) = 0.5833333333333334

|  | **X=1** | **X=2** | **X=3** | **X=4** | **X=5** | **X=6** |
| --- | --- | --- | --- | --- | --- | --- |
| **Y=6** |  |  |  |  |  | 0.027778 |
| **Y=5** |  |  |  |  | 0.055556 |  |
| **Y=4** |  |  |  | 0.083333 |  |  |
| **Y=3** |  |  | 0.111111 |  |  |  |
| **Y=2** |  | 0.138889 |  |  |  |  |
| **Y=1** | 0.166667 |  |  |  |  |  |

计算，代码方法如下。

joint\_dist.event(lambda i,j:(pow(i,2)+pow(j,2))<8, 'X', 'Y')

P(Event) = 0.19444444444444442

|  | **X=1** | **X=2** | **X=3** | **X=4** | **X=5** | **X=6** |
| --- | --- | --- | --- | --- | --- | --- |
| **Y=6** |  |  |  |  |  |  |
| **Y=5** |  |  |  |  |  |  |
| **Y=4** |  |  |  |  |  |  |
| **Y=3** |  |  |  |  |  |  |
| **Y=2** | 0.0 |  |  |  |  |  |
| **Y=1** | 0.166667 | 0.027778 |  |  |  |  |

* 二维连续型随机变量及其联合密度函数

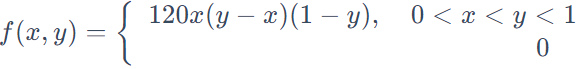
定义 7：设二维随机变量的联合分布函数为，如果存在一个二元非负实值函数，使得对于任意有成立，则称为**二维连续型随机变量**，为二维连续型随机变量的**联合（概率）密度函数**。式中，表示二重积分，其中积分区域。

定义 8：设维随机变量的联合分布函数为，如果存在一个元非负函数，使得对任意的有$F(x\_1, x\_2, , x\_n)=*{-}^{x\_1}* {-}^{x\_n} f(u\_1, u\_2, , u\_n) u\_1 u\_2 u\_n $ 成立，则称为**维连续型随机变量**，为维连续型随机变量的**联合（概率）密度函数**。

二维连续型随机变量的联合密度函数有：

定理 2（**联合密度函数的性质**）：设为二维连续型随机变量的联合密度函数，则

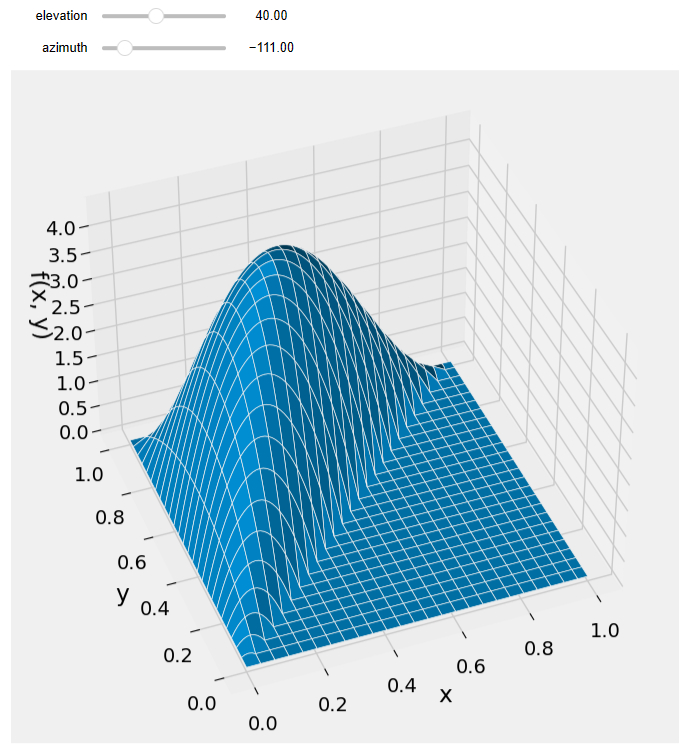
1. 非负性，;
2. 规范性 。

联合密度函数的规范性意味着以曲面为顶，以整个平面与的交集区域为底的曲顶柱体的体积为1。如下图，其联合密度函数为否则 。

使用[SimPy](https://docs.sympy.org/latest/modules/integrals/integrals.html)②和[datascience](https://github.com/data-8/datascience)①库定义二维连续型随机变量的联合密度函数并打印显示。

from sympy import Symbol,Integral  
import numpy as np   
import matplotlib.pyplot as plt

def joint(x,y):  
 if y < x:  
 return 0  
 else:  
 return 120 \* x \* (y-x) \* (1-y)  
   
Plot\_3d(x\_limits=(0,1), y\_limits=(0,1), f=joint, cstride=4, rstride=4,interactive=True)



x = Symbol('x', positive=True)  
y = Symbol('y', positive=True)  
  
f = 120\*x\*(y-x)\*(1-y)  
f

使用Integral方法定义二重积分。

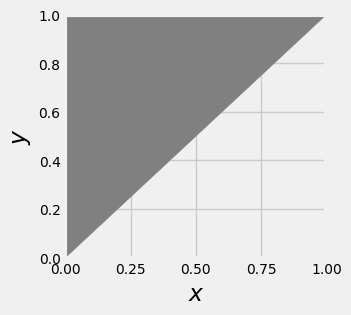
F\_XY=Integral(f, (x, 0, y), (y, 0, 1))  
F\_XY

doit()方法估计该二重积分的结果为1。

F\_XY.doit()

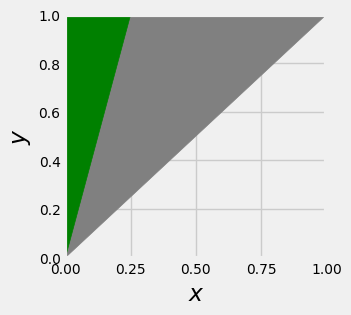
为了方便观察随机变量的数值区间，以平面形式表述。

coords=np.arange(0,1+0.1,0.1)  
pts\_y=[[0,i] for i in coords]  
pts\_x=[[i,i] for i in coords]  
pts\_x.reverse()  
pts=np.array(pts\_y+pts\_x)  
  
plt.figure(figsize=(3,3))  
t=plt.Polygon(pts, color='gray')  
plt.gca().add\_patch(t)  
  
plt.xlabel("$x$")  
plt.ylabel("$y$")  
plt.xticks(fontsize=10);plt.yticks(fontsize=10)  
plt.show()



计算概率，代码计算如下。

pts\_x\_d1=[[i/4,i] for i in coords]  
pts\_x\_d1.reverse()  
pts\_d1=np.array(pts\_y+pts\_x\_d1)  
  
plt.figure(figsize=(3,3))  
t=plt.Polygon(pts, color='gray')  
plt.gca().add\_patch(t)  
  
t1=plt.Polygon(pts\_d1,color='g')  
plt.gca().add\_patch(t1)  
  
plt.xlabel("$x$")  
plt.ylabel("$y$")  
plt.xticks(fontsize=10);plt.yticks(fontsize=10)  
plt.show()

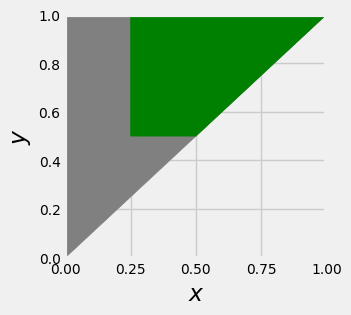


f\_1=Integral(f, (x, 0, y/4), (y, 0, 1))  
f\_1

f\_1.doit()

计算概率，代码计算如下。

coords\_d2=[i for i in coords if i>=0.5]  
pts\_y\_d2=[[0.25,i] for i in coords\_d2]  
  
pts\_x\_d2=[[i,i] for i in coords\_d2 if i >=0.25]  
pts\_x\_d2.reverse()  
pts\_d2=np.array(pts\_y\_d2+pts\_x\_d2)  
  
plt.figure(figsize=(3,3))  
t=plt.Polygon(pts, color='gray')  
plt.gca().add\_patch(t)  
  
t2=plt.Polygon(pts\_d2,color='g')  
plt.gca().add\_patch(t2)  
  
plt.xlabel("$x$")  
plt.ylabel("$y$")  
plt.xticks(fontsize=10);plt.yticks(fontsize=10)  
plt.show()



f\_2=Integral(f, (x, 0.25, y), (y, 0.5, 1))  
f\_2

f\_2.doit()

* 常用的多维随机变量

1）二维均匀分布

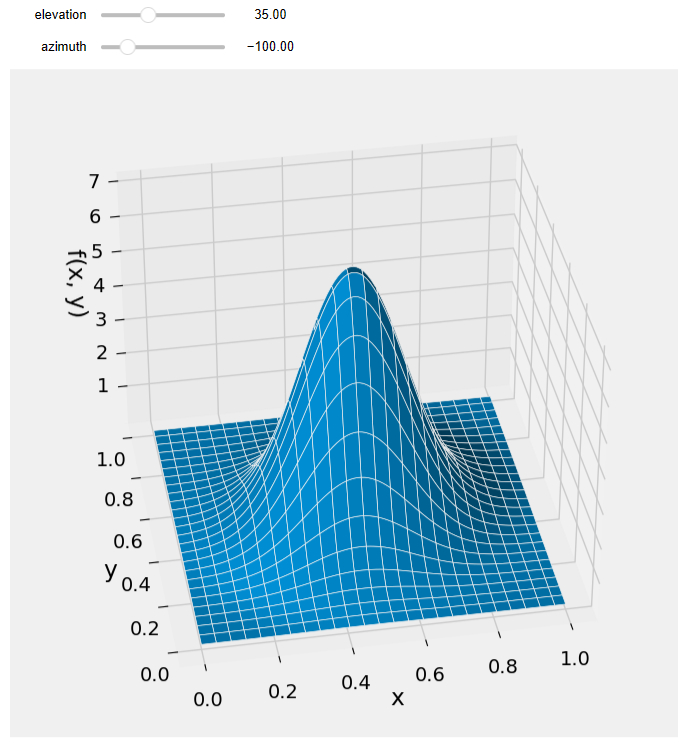
设二维随机变量的联合密度函数为，式中，是平面上的某个区域，为的面积，则称服从区域上的**二维均匀分布**。

2）二维正态分布

如果的联合密度函数为，则称服从**二维正态分布**，并记为，式中，−*∞<*μ*1​,*μ*2​<+∞,*σ*1​*σ*2​>0,∣*ρ*∣<1*。

二维正态分布的联合密度函数图像如下。

def joint(x,y,mu1=0.5,mu2=0.5,sig1=0.15,sig2=0.15,rho=0.1):  
 return 1/(2\*np.pi\*sig1\*sig2\*np.sqrt(1-rho\*\*2))\*np.exp(-1/(2\*(1-rho\*\*2))  
 \*(((x-mu1)\*\*2/sig1\*\*2)   
 - 2\*rho\*(x-mu1)\*(y-mu2)/(sig1\*sig2)   
 + (y-mu2)\*\*2/sig2\*\*2))  
   
Plot\_3d(x\_limits=(0,1), y\_limits=(0,1), f=joint, cstride=4, rstride=4,interactive=True)



* 边缘分布

二维随机变量中，和也都是随机变量，分别有各自的概率分布，称和的概率分布为二维随机变量关于和关于的**边缘概率分布**，简称**边缘分布**。

定义 1：设二维随机变量的联合分布函数为，称为**随机变量的边缘分布函数**；称P(Y⩽y)=P(X⩽+∞,Y⩽y)=F(+∞,y),−∞<y<+∞Y**的边缘分布函数**。

定义 2：设二维离散型随机变量的联合分布律为，称概率P(X=xi​)=P(X=xi​,∪j​Y=yj​)=∑j​P(X=xi​,Y=yj​)=∑j​pij​,i=1,2,⋯*, Xpi .*pi​.=P(X=xi​)=∑j​pij​,i=1,2,⋯ *P(Y=y\_j), j=1,2, YPj *X(X,Y)Y(X,Y)联合分布律表格中的行（或列）和。因为边缘分布律位于联合分布律表格的边缘，因此称其为边缘分布律。

下面计算了前文关于一个年级数学和语文成绩分布的联合分布表及其边缘分布律。

from sympy import Symbol,Integral  
import numpy as np   
import matplotlib.pyplot as plt  
import matplotlib  
import pandas as pd  
from fractions import Fraction

k=np.arange(2)  
  
def joint\_probability(x,y):  
 if x==0 and y==0: return 0.78  
 elif x==0 and y==1: return 0.02  
 elif x==1 and y==0: return 0.12  
 elif x==1 and y==1: return 0.08  
  
joint\_table=Table().values('X',k,'Y',k).probability\_function(joint\_probability)  
joint\_table.both\_marginals()

|  | **X=0** | **X=1** | **Sum: Marginal of Y** |
| --- | --- | --- | --- |
| **Y=1** | 0.02 | 0.08 | 0.1 |
| **Y=0** | 0.78 | 0.12 | 0.9 |
| **Sum: Marginal of X** | 0.80 | 0.20 | 1.0 |

joint\_table.marginal('Y')

|  | **X=0** | **X=1** | **Sum: Marginal of Y** |
| --- | --- | --- | --- |
| **Y=1** | 0.02 | 0.08 | 0.1 |
| **Y=0** | 0.78 | 0.12 | 0.9 |

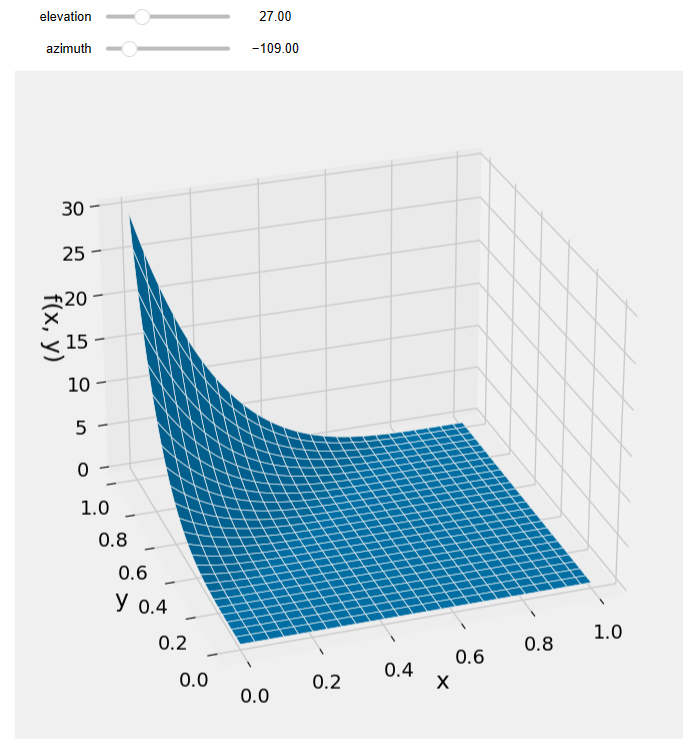
joint\_table.marginal('X')

|  | **X=0** | **X=1** |
| --- | --- | --- |
| **Y=1** | 0.02 | 0.08 |
| **Y=0** | 0.78 | 0.12 |
| **Sum: Marginal of X** | 0.80 | 0.20 |

定义 3：设二维连续型随机变量的联合分布函数为，联合密度函数为，根据，得，所以**的边缘密度函数**为；类似，**的边缘密度函数**为。

例如，定义二维连续型随机变量的联合密度函数，绘制图形如下。

def jt\_dens(x,y):  
 if y < x:  
 return 0  
 else:  
 return 30 \* (y-x)\*\*4  
  
Plot\_3d(x\_limits=(0,1), y\_limits=(0,1), f=jt\_dens, cstride=4, rstride=4,interactive=True)



联合密度函数的二重积分定义为变量jt\_dens\_pdf。

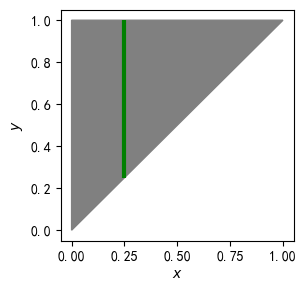
x = Symbol('x', positive=True)  
y = Symbol('y', positive=True)  
  
joint\_density = 30\*(y-x)\*\*4  
jt\_dens\_pdf=Integral(joint\_density, (y, x, 1), (x, 0, 1))  
jt\_dens\_pdf

估计二重积分的结果为1。

jt\_dens\_pdf.doit()

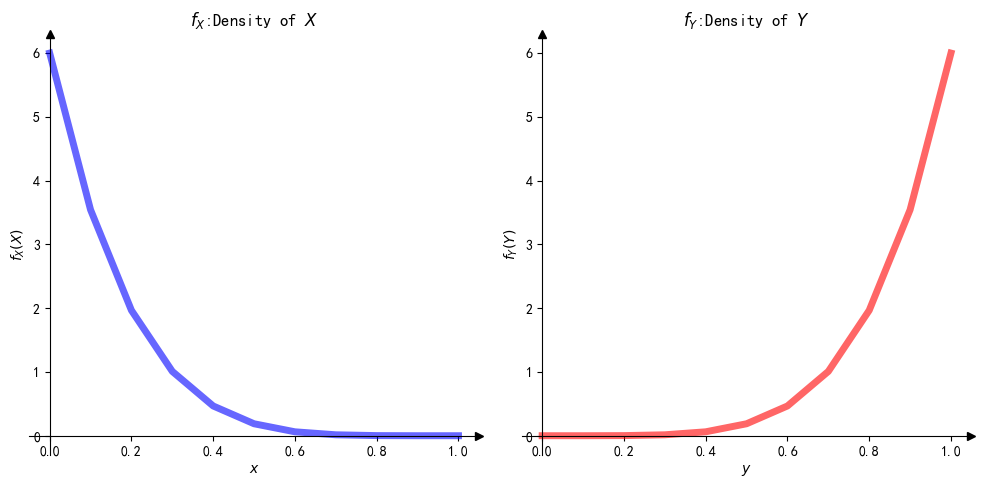
下图表述了约为0.25时，事件的概率，为固定，沿的积分。

coords=np.arange(0,1+0.1,0.1)  
pts\_y=[[0,i] for i in coords]  
pts\_x=[[i,i] for i in coords]  
pts\_x.reverse()  
pts=np.array(pts\_y+pts\_x)  
  
plt.figure(figsize=(3,3))  
t=plt.Polygon(pts, color='gray')  
plt.gca().add\_patch(t)  
  
plt.vlines(x=0.25,ymin=0.25,ymax =1, linewidth=3, color='green',linestyles='-')  
  
plt.xlabel("$x$")  
plt.ylabel("$y$")  
plt.xticks(fontsize=10);plt.yticks(fontsize=10)  
plt.show()



计算的边缘密度函数；计算的边缘密度函数。下面打印了两者的边缘密度函数曲线。

import matplotlib as mpl  
mpl.rcParams.update(mpl.rcParamsDefault)  
  
def marginal\_density\_X(x):  
 return 6\*(1-x)\*\*5  
  
def marginal\_density\_Y(y):  
 return 6\*y\*\*5  
  
plt.rcParams['font.sans-serif'] = ['SimHei']  
matplotlib.rcParams['axes.unicode\_minus'] = False  
  
fig, axs=plt.subplots(1, 2,figsize=(10,5))  
x=np.arange(0,1+0.1,0.1)  
  
axs[0].plot(x, marginal\_density\_X(x),'b-', lw=5, alpha=0.6)  
axs[1].plot(x, marginal\_density\_Y(x),'r-', lw=5, alpha=0.6)  
  
axs[0].set\_title("$f\_{X}$:Density of $X$")  
axs[1].set\_title("$f\_{Y}$:Density of $Y$")  
  
  
axs[0].set\_xlabel("$x$")  
axs[0].set\_ylabel("$f\_{X}(X)$")  
axs[1].set\_xlabel("$y$")  
axs[1].set\_ylabel("$f\_{Y}(Y)$")  
plot\_style(axs[0])  
plot\_style(axs[1])  
fig.tight\_layout()  
plt.show()



定理 1：如果(X,Y)∼N(μ1​,μ2​,σ12​,σ22​,ρ) X∼N(μ1​,σ12​),Y∼N(μ2​,σ22​)。

* 条件分布

对于二维随机变量中的两个随机变量和，在许多问题中它们的取值往往彼此影响，这使得条件分布成为研究变量之间相依关系的有利工具。对于二维随机变量，随机变量的条件分布是在给定取某个值条件下的分布。

1）二维离散型随机变量的条件分布律

定义 1：设二维离散型随机变量的联合分布律为，对于固定的，记在给定条件下的随机变量为，其值域记为。条件分布律满足分布律的两条性质：

（1）；

（2）.

当时，在给定条件下的随机变量的条件分布律为$P(Y=y\_j X=x\_i)=, j=1,2, x\_i *X{X=x\_i}YY X=x\_i* (xi.. 固定 。同理，条件分布律也满足分布律的两条性质。

定义一个二维离散型随机变量的条件分布律表格joint\_table，如下。

k = np.arange(3)  
arr=np.array([[Fraction(1,18),Fraction(2,9),Fraction(1,18)],[Fraction(2,15),Fraction(2,15),Fraction(1,15)],[Fraction(1,6),Fraction(1,9),Fraction(1,18)]])  
print(pd.DataFrame(arr))  
  
def joint\_probability(x, y):  
 return (arr[x,y])  
  
joint\_table=Table().values('X', k, 'Y', k).probability\_function(joint\_probability)  
print(f'total probability={joint\_dist.total\_probability()}')  
joint\_table

0 1 2  
0 1/18 2/9 1/18  
1 2/15 2/15 1/15  
2 1/6 1/9 1/18  
total probability=1.0

|  | **X=0** | **X=1** | **X=2** |
| --- | --- | --- | --- |
| **Y=2** | 0.055556 | 0.066667 | 0.055556 |
| **Y=1** | 0.222222 | 0.133333 | 0.111111 |
| **Y=0** | 0.055556 | 0.133333 | 0.166667 |

计算随机变量和的边缘分布律。

joint\_table.both\_marginals()

|  | **X=0** | **X=1** | **X=2** | **Sum: Marginal of Y** |
| --- | --- | --- | --- | --- |
| **Y=2** | 0.055556 | 0.066667 | 0.055556 | 0.177778 |
| **Y=1** | 0.222222 | 0.133333 | 0.111111 | 0.466667 |
| **Y=0** | 0.055556 | 0.133333 | 0.166667 | 0.355556 |
| **Sum: Marginal of X** | 0.333333 | 0.333333 | 0.333333 | 1.000000 |

假设知道，则提取满足条件的行，可知该行和的概率为，并不为1。由，计算给定条件时的条件分布为，。可以直接调用conditional\_dist方法计算。

def indicator\_Y\_equals\_1(i, j):  
 return j==1  
  
joint\_table.event(indicator\_Y\_equals\_1, 'X', 'Y')

P(Event) = 0.4666666666666666

|  | **X=0** | **X=1** | **X=2** |
| --- | --- | --- | --- |
| **Y=2** |  |  |  |
| **Y=1** | 0.222222 | 0.133333 | 0.111111 |
| **Y=0** |  |  |  |

以为条件计算的条件分布律。

joint\_table.conditional\_dist('X', 'Y') # conditional distribution of X given each different value of Y

|  | **X=0** | **X=1** | **X=2** | **Sum** |
| --- | --- | --- | --- | --- |
| **Dist. of X | Y=2** | 0.312500 | 0.375000 | 0.312500 | 1.0 |
| **Dist. of X | Y=1** | 0.476190 | 0.285714 | 0.238095 | 1.0 |
| **Dist. of X | Y=0** | 0.156250 | 0.375000 | 0.468750 | 1.0 |
| **Marginal of X** | 0.333333 | 0.333333 | 0.333333 | 1.0 |

以为条件计算的条件分布律。

joint\_table.conditional\_dist('Y', 'X') # conditional distribution of Y given each different value of X

|  | **Dist. of Y | X=0** | **Dist. of Y | X=1** | **Dist. of Y | X=2** | **Marginal of Y** |
| --- | --- | --- | --- | --- |
| **Y=2** | 0.166667 | 0.2 | 0.166667 | 0.177778 |
| **Y=1** | 0.666667 | 0.4 | 0.333333 | 0.466667 |
| **Y=0** | 0.166667 | 0.4 | 0.500000 | 0.355556 |
| **Sum** | 1.000000 | 1.0 | 1.000000 | 1.000000 |

2）二维连续型随机变量的条件密度函数

定义 2：设为二维连续型随机变量的联合密度函数，当时，在给定条件下的条件密度函数为。对于固定的，记在给定条件下的随机变量为，其值域记为 固定 。条件密度函数满足密度函数的两条性质：

（1）；

（2）。

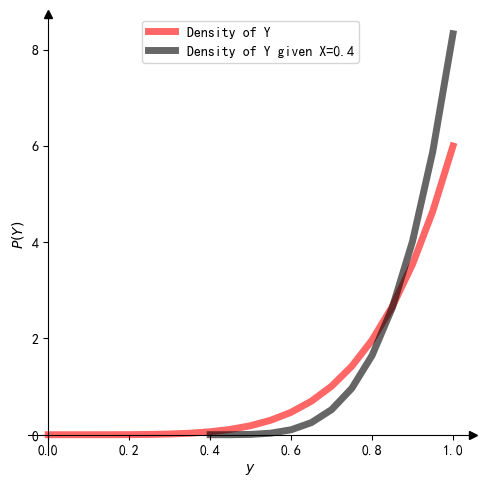
当时，在给定条件下的条件密度函数为。对于固定的，记在给定条件下的随机变量为，其值域记为 固定 。同理，可以验证满足密度函数的两条性质。

定义 3：设为二维连续型随机变量的联合密度函数，当时，在给定条件下的条件分布函数为；当时，在给定条件下的条件分布函数为, 其中 。

继续*边缘分布：二维连续型随机变量的联合密度函数*部分的案例，给定条件时，的条件密度函数为。同时绘制的边缘密度函数和条件密度函数，如下。

y=Symbol('y', positive=True)  
conditional\_density\_Y\_given\_X\_is\_04=(5/(0.6\*\*5)) \* (y - 0.4)\*\*4  
Integral(conditional\_density\_Y\_given\_X\_is\_04, (y, 0.4, 1)).doit()

def conditional\_density\_Y\_given\_X\_is\_04\_func(y):  
 return 5/0.6\*\*5\*(y-0.4)\*\*4  
  
def marginal\_density\_Y(y):  
 return 6\*y\*\*5  
  
fig, ax=plt.subplots(1, 1,figsize=(5,5))  
x=np.arange(0,1+0.05,0.05)  
x\_c=np.array([i for i in x if i>=0.4])  
  
ax.plot(x, marginal\_density\_Y(x),'r-', lw=5, alpha=0.6,label="Density of Y")  
ax.plot(x\_c, conditional\_density\_Y\_given\_X\_is\_04\_func(x\_c),'k-', lw=5, alpha=0.6,label="Density of Y given X=0.4")  
  
ax.set\_xlabel("$y$")  
ax.set\_ylabel("$P(Y)$")  
plot\_style(ax)  
ax.legend(loc='upper center')  
fig.tight\_layout()  
plt.show()



由条件密度函数可以计算条件概率（probabilities）和条件期望（expectations），例如。

p\_1=Integral(conditional\_density\_Y\_given\_X\_is\_04, (y, 0.9, 1))  
print(p\_1.doit())  
p\_1

0.598122427983537

由条件密度函数计算条件期望值，。

Integral(y\*conditional\_density\_Y\_given\_X\_is\_04, (y, 0.4, 1)).doit()

* 随机变量的独立性

当两个随机变量的取值规律互不影响时，称它们是相互独立的。

定义 4：设为二维随机变量，若对任意，都有成立，则称**随机变量与相互独立**。式中，为的联合分布函数，和FY(y)XY的边缘分布函数。

定理 2：设计为二维离散型随机变量，那么，与相互独立的充分必要条件是对任意的，都有成立，式中，为的联合分布律，和分别为和的边缘分布律。相互独立性的直观含义是当取定时，的取值规律不受任何影响，即

例如下述二维随机变量中和相互独立，可以验证都有成立。

k=np.arange(2)  
  
def joint\_probability(x,y):  
 if x==0:return 0.4  
 if x==1: return 0.1  
  
joint\_table=Table().values('X',k,'Y',k).probability\_function(joint\_probability)  
joint\_table.both\_marginals()

|  | **X=0** | **X=1** | **Sum: Marginal of Y** |
| --- | --- | --- | --- |
| **Y=1** | 0.4 | 0.1 | 0.5 |
| **Y=0** | 0.4 | 0.1 | 0.5 |
| **Sum: Marginal of X** | 0.8 | 0.2 | 1.0 |

定理 3：若为二维连续型随机变量，那么，与相互独立的充分必要条件是在、及的一切公共连续点上都有成立。式中，为的联合密度函数， fX(x) fY(y)XY的边缘密度函数。

定义 5：设为维随机变量，若对任意，都有 X1, X2, , XnF(x1, x2, , xn)(X1, X2, , Xn)   
FXi​​(xi​)Xii=1,2,… ,n。

当为离散型随机变量时，随机变量相互独立的充要条件是对任意的，都有，成立，式中，为的联合分布律，为的边缘分布律，。

当为连续型随机变量时，随机变量相互独立的充要条件是在的一切公共连续点上都有成立。式中，为的联合密度函数，为的边缘密度函数，。

* 二维随机变量函数的分布

设为二维随机变量，则是的函数，且是一维随机变量。如果要由的分布求的分布，一般要将的分布转化成有关的概率分布。

1）二维离散型随机变量函数的分布

设是二维离散型随机变量，其分布律为，是的函数，则也是离散型随机变量，其分布律为。

定理 1：设，且与相互独立，则；设，且与相互独立，则。

2）二维连续型随机变量函数的分布

设二维连续型随机变量的联合密度函数为，则随机变量的二元函数的分布函数为，式中，是与等价的随机事件，而是平面上的点集（通常是一个区域或若干个区域的并集）。则的密度函数为。这种计算二维连续型随机变量函数分布的方法称为分布函数法。

定理 2：设随机变量的联合密度函数为，且的边缘密度函数为，的边缘密度函数为，则随机变量的函数的密度函数为。特别地，当随机变量和相互独立时，。这两个公式称为**卷积公式**。在概率论中计算相互独立随机变量之和分布的运算称为卷积运算。

定理 3：设，且与相互独立，则。

定理 4：设连续型随机变量和相互独立，且的分布函数为，的分布函数为，则

（1）随机变量的分布函数为；

（2）随机变量的分布函数为。

由定理 4，推广至个相互独立随机变量的情形有，设连续型随机变量相互独立，且的分布函数为，则

（1）随机变量的分布函数为；

（2）随机变量的分布函数为。

## 3.5.4 随机变量的数字特征

* 级数（补充）[4]

**级数**（Series）是数学中一个有穷或无穷的序列，例如之和，即，如果序列是有穷序列，其和称为有穷级数；反之，称为无穷级数（一般也简称为级数）。序列中的项称作级数的通项（或一般项）。级数的通向可以是实数、矩阵或向量等常量，也可以是关于其它变量的函数，不一定是一个数。一般的，如果级数的通项是常量，则称之为常数项级数，如果级数的通项是函数，则称之为函数项级数。常见简单有穷数量的级数有等差数列和等比数列的级数。

无穷级数有发散和收敛的区别，称为无穷级数的敛散性。无穷级数在收敛时才会有一个和；发散的无穷级数在一般意义上没有和，但可以用一些别的方式来定义。

无穷级数一般写作、或。级数收敛时，其和通常被表示为。

**无穷级数的定义**，设是一个无穷序列：，其前项的和称为的**部分和**：。部分和依次构成另一个无穷序列：，这两个序列合成为一个级数，记作或。

**无穷级数的敛散性**，对于级数，如果当趋于正无穷大时，趋向一个有限的极限：，那么这个无穷数就是收敛的，称为的和。如果极限不存在，这个无穷级数就是发散的。收敛的无穷级数存在唯一的一个和。这时可以定义级数的余项和：。

**任意项级数**，如果级数中的各项可以是正数、负数或零，则级数称为任意项级数。将任意项级数各项取绝对值，得到正项级数，。

**条件收敛**，如果任意项级数收敛，而级数发散，则称级数条件收敛。

**绝对收敛**，如果级数收敛，则称级数绝对收敛。

定理：如果任意性级数各项绝对值所组成的正向级数收敛，则级数收敛。

收敛级数的性质：

* 若一个无穷级数收敛，其和为，则如果每一项乘以一个常数，得到的级数也收敛，且和等于。
* 收敛的无穷级数可以逐项相加或相减，如有两个无穷级数：和，则。
* 级数前面加上有限项或减去有限项不影响其收敛性，如：和，这两个级数的敛散性是一样的。
* 当趋向无限大时，任何一个收敛级数的通项都趋于0：。

借助[SymPy](https://docs.sympy.org/latest/modules/concrete.html)②级数求和，例如，

from IPython.display import display, Latex  
from sympy import I, oo, Sum, exp, pi,factorial, S,Symbol, oo,pprint,print\_latex,pi,Integral,Product  
from sympy.abc import x, y, a, b, c, d, n,p,i  
import numpy as np

Series\_Sum\_1=Sum(x\*\*n/factorial(n),(n,0,oo))  
Series\_Sum\_1

Series\_Sum\_1.doit()

Series\_Sum\_2=Sum(n\*p\*\*(n-1),(n,1,oo))  
Series\_Sum\_2

Series\_Sum\_2.doit()

Series\_Sum\_3=Sum(p\*\*(n+1)/(n+1),(n,0,oo))  
Series\_Sum\_3

Series\_Sum\_3.doit()

用is\_convergent方法检查级数的敛散性。is\_convergent检验求和的（Sum）收敛性；is\_absolutely\_convergent检验无穷级数的绝对收敛性。

Series\_Sum\_4=Sum(1/n\*\*(S(6)/5), (n, 1, oo))  
print(f'is convergent:{Series\_Sum\_4.is\_convergent()}')  
print(f'is absolutely convergent:{Series\_Sum\_4.is\_absolutely\_convergent()}')  
Series\_Sum\_4

is convergent:True  
is absolutely convergent:True

* 数学期望

定义 1：设是离散型的随机变量，其分布律为，如果级数绝对收敛，则称为**离散型随机变量的数学期望**，也称作**期望**或**均值**。

定义中要求级数绝对收敛，是为了保证数学期望的唯一性。如级数条件收敛，级数改变项的次序后，其和不唯一。只有当绝对收敛时，改变项的顺序才不影响和的唯一性，即绝对收敛级数具有可交换性。

例如，有随机变量的分布律为，，检查的敛散性为False，即不存在。

ev\_1=Sum((2\*\*i/i)\*(1/2\*\*i),(i,1,oo))  
print(ev\_1.is\_absolutely\_convergent())  
ev\_1

False

同样，对于，不存在。

ev\_2=Sum(((-1)\*\*i\*2\*\*i/i)\*(1/2\*\*i),(i,1,oo))  
print(ev\_2.is\_absolutely\_convergent())  
ev\_2

False

对于，则存在，为。

ev\_3=Sum(((-1)\*\*i\*2\*\*i/i\*\*2)\*(1/2\*\*i),(i,1,oo))  
print(ev\_3.is\_absolutely\_convergent())  
ev\_3

True

ev\_3\_result=ev\_3.doit()  
print\_latex(ev\_3\_result)  
ev\_3\_result

- \frac{\pi^{2}}{12}

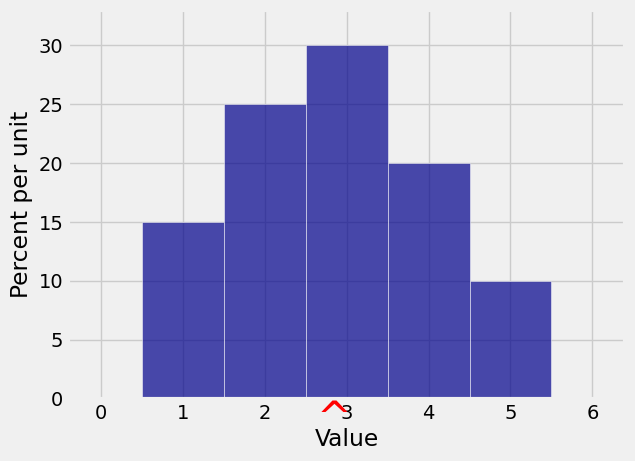
又例如，随机变量的分布律px\_dist如下定义，计算期望值为2.85。

x=np.arange(1, 6)  
probs=make\_array(0.15, 0.25, 0.3, 0.2, 0.1)  
px\_dist=Table().values(x).probabilities(probs)  
px\_dist

| **Value** | **Probability** |
| --- | --- |
| 1 | 0.15 |
| 2 | 0.25 |
| 3 | 0.3 |
| 4 | 0.2 |
| 5 | 0.1 |

display(Latex(f'$E(X)$={px\_table.ev()}')) # 同 sum(x\*probs)  
Plot(px\_dist,show\_ev=True)

=2.8500000000000005



离散型随机变量数学期望的定义可推广至连续型随机变量的情形。首先将连续性随机变量的取值“离散化”，即将的值域分割成份，记第份为，在每一个小区间，的取值近似为，相应的概率为。“离散化”后视为离散型随机变量，其分布律近似为，

|  |  |
| --- | --- |
| 概率 |  |

则的平均值近似为，另个小区间长度的最大值，得到的平均值精确为，这就是连续型随机变量的数学期望。和离散型随机变量类似，要求绝对收敛。

定义 2：设是连续型随机变量，其密度函数为。如果广义积分绝对收敛，则称为**连续型随机变量的数学期望**，也称作**期望或均值**。

例如，设随机变量的密度函数为，由下述计算可知其不存在。

# 定义密度函数  
density\_1=1/pi\*1/(1+x\*\*2)  
density\_1

# 计算围合的总面积，满足规范性  
total\_area\_1=Integral(density\_1,(x,-oo,+oo))  
print(f'total area={total\_area\_1.doit()}')  
total\_area\_1

total area=1

通过计算期望，结果为nan，因此对于该案例的随机变量，并不存在。

# 定义期望  
expectation\_1=Integral(x\*density\_1, (x, -oo,+oo))  
print(expectation\_1.doit())  
expectation\_1

Nan

* 随机变量函数的数学期望

定理 1（**随机变量一元函数的期望公式**）：

（1）设是离散型随机变量，其分布律为，如果级数绝对收敛，则的一元函数的数学期望为。

（2）设是连续型随机变量，其密度函数为，如果广义积分绝对收敛，则的一元函数的数学期望为。

定理 2（**随机变量二元函数的期望公式**）：

（1）设是二维离散型随机变量，其联合分布律为，如果级数绝对收敛，则二元函数的数学期望为；特别地，。

（2）设是二维连续型随机变量，其联合密度为。如果广义积分绝对收敛，则的二元函数的数学期望为；特别地，，。

* 数学期望的性质

（1）设为常数，则；

（2）设为随机变量，且存在，为常数，则；

（3）设为任意两个随机变量，且和存在，则；

（4）设与为相互独立的随机变量，且和存在，则。

* 方差和标准差

定义 设是一个随机变量，如果存在，则称为**随机变量的方差**。称方差的算数平方根为**随机变量的标准差**。实际计算方差时，可用计算。

定理 方差的性质

（1）的充分必要条件是，即服从参数为的退化分布，其中。特别，若为常数，则;

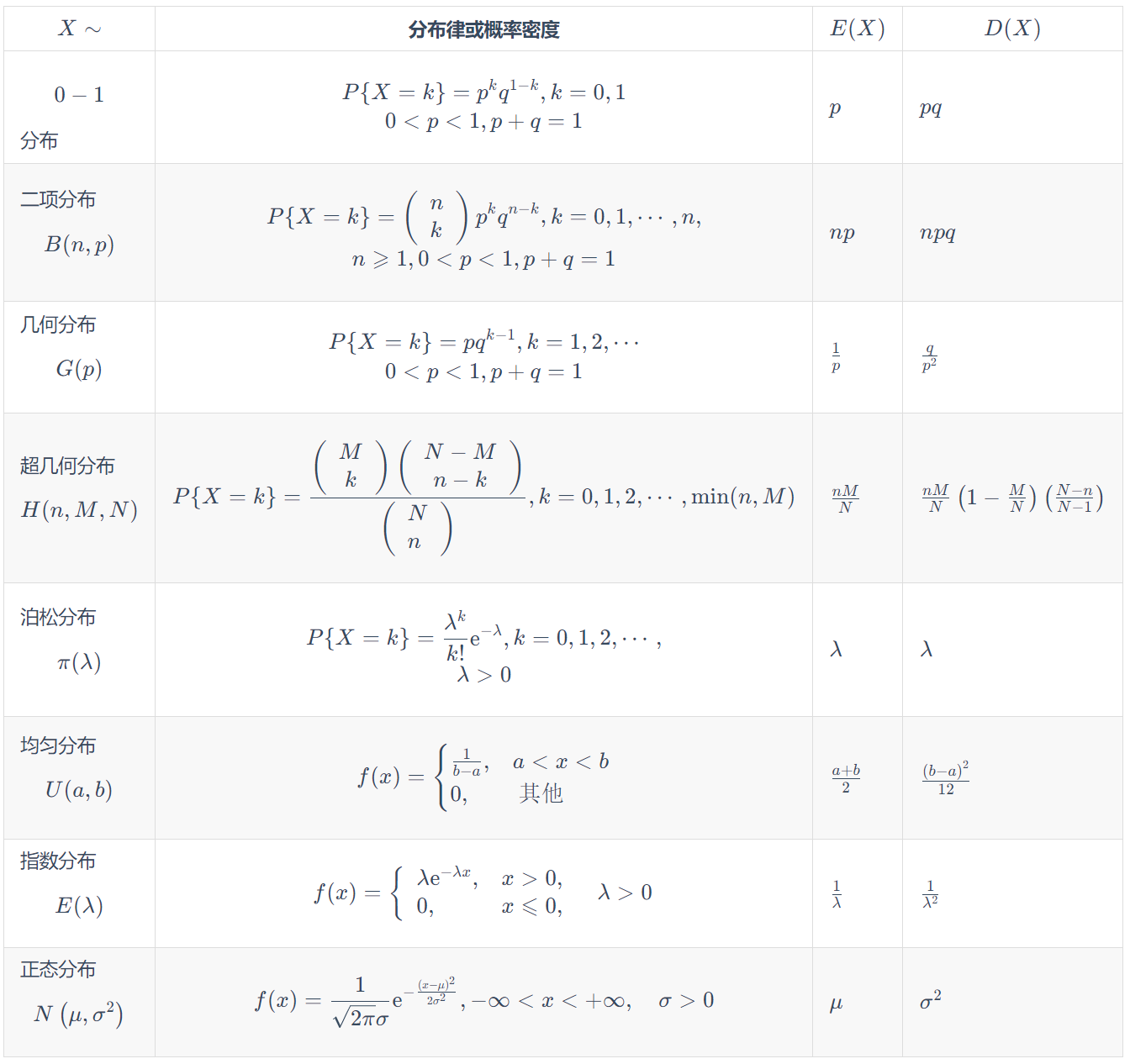
（2）设为随机变量，为常数，则；

（3）设为相互独立的随机变量，则；

（4）设与为相互独立的随机变量，则。

* 常用分布及其数学期望与方差

| X $ | 分布律或概率密度 |  |  |
| --- | --- | --- | --- |
| 分布 | $$P\{X=k\}=p^k q^{1-k}, k=0,1 \\0<p<1, p+q=1$$ |  |  |
| 二项分布 | $$P\{X=k\}=\left(\begin{array}{l}n \\k\end{array}\right) p^k q^{n-k}, k=0,1, \cdots, n, \\n \geqslant 1,0<p<1, p+q=1$$ |  |  |
| 几何分布 | $$P\{X=k\}=p q^{k-1}, k=1,2, \cdots \\0<p<1, p+q=1$$ |  |  |
| 超几何分布 |  |  | $(1-)() $ |
| 泊松分布 | $$P\{X=k\}=\frac{\lambda^k}{k !} \mathrm{e}^{-\lambda}, k=0,1,2, \cdots, \\\lambda>0$$ |  |  |
| 均匀分布 |  |  |  |
| 指数分布 |  |  |  |
| 正态分布 |  | $$ | $ ^2$ |
|  |  |  |  |



* 切比雪夫不等式

定理 （**切比雪夫不等式**），设随机变量的数学期望与方差存在，则对于任意正数，不等式成立。切比雪夫不等式可书写为如下形式，。这个不等式给出了在随机变量的分布未知情况下，事件概率的下限估计。

* 协方差和相关系数

定义 1 设是二维随机变量，如果存在，则称为**随机变量和的协方差**。实际计算协方差时，更多使用如下公式，。

定理 1（协方差的性质），设为任意的随机变量，为常数，则有，

（1）；

（2）；

（3）；

（4）$(X\_1+X\_2, Y)=(X\_1, Y)+(X\_2, Y) $。

协方差考察了随机变量之间协同变化的关系，但为了避免量纲的影响，将随机变量标准化，，再求协方差，就是随机变量和的相关系数，又称为标准化协方差。因为，所以定义相关系数如下。

定义 2，设是二维随机变量，如果存在，且，则称为**随机变量和的相关系数**，也记作。

定义 3，设是二维随机变量，当时，称与（线性）无关或（线性）不相关。

定理 2，当时，下列5个命题是等价的：（1）；（2）；（3）；（4）；（5）。

定理 2，相关系数的性质，设是二维随机变量，当存在且时，有

（1）；

（2）的充要条件是，其中，

当时，,

当时，；

（3）若随机变量与相互独立，则与线性无关，即。但由不能推断与相互独立。

定义 4，设二维随机变量的相关系数存在，则

当时，的取值在直线上的概率为1，称与完全线性相关；

当时，的取值在斜率大于0的直线上的概率为1，称与完全正线性相关；

当时，的取值在斜率小于0的直线上的概率为1，称与完全负线性相关；

当时，称与正线性相关；

当时，称与负线性相关。

定理 4，如果二维随机变量服从二维正态分布，则与相互独立等价于与不相关。

注释（Notes）：

① datascience，（<https://github.com/data-8/datascience>）。

② SymPy，（<https://docs.sympy.org/latest/index.html>）。

参考文献（References）:

[1] (工业和信息化“十二五”规划教材) 同济大学数学系. 概率论与数理统计[M].人民邮电出版社 (2017)

[2] 主编：许伯生 刘春燕；参编：刘瑞娟 肖翔 朱萌 洪银萍 周宇. 概率论与数理统计（第2版）[M].清华大学出版社 (2019)

[3] 数据科学的概率课程（Probability for Data Science）在线教材，<http://prob140.org/textbook/content/README.html>.

[4] 级数（Wikipedia），<https://en.wikipedia.org/wiki/Series_(mathematics)><https://zh.wikipedia.org/wiki/%E7%BA%A7%E6%95%B0>.