概率论

✓ 概率论是干什么的呢?

研究随机现象数量规矩的数学分支。

概率论

- 随机事件是什么呢?

- 扔硬币，王者峡谷击杀数，一批产品合格数。。。这些有什么特点呢?

- 1.可以在相同条件下重复执行
  2.事先就能知道可能出现的结果
  3.试验开始前并不知道这一次的结果

- 随机试验E的所有结果构成的集合称为E的样本空间 $S=\{e\}$
  抛硬币:$S=\{$正面，反面$\}$
  击杀数:$S=\{0,1,2,\ldots\}$

频率与概率

A在这N次试验中发生频率：$f_n(A)=\dfrac{n_A}{n}$

其中，n_A—A发生的次数(频数)；n—总试验次数。

$f_n(A)$ 的稳定值P定义为A的概率$\boldsymbol{P(A)=p}$

概率与频率

试验序号	$n=5$		$n=50$		$n=500$	
	n_H	$f_n(H)$	n_H	$f_n(H)$	n_H	$f_n(H)$
1	2	0.4	22	0.44	251	0.502
2	3	0.6	25	0.50	249	0.498
3	1	0.2	21	0.42	256	0.512
4	5	1.0	25	0.50	253	0.506
5	1	0.2	24	0.48	251	0.502
6	2	0.4	21	0.42	246	0.492
7	4	0.8	18	0.36	244	0.488
8	2	0.4	24	0.48	258	0.516
9	3	0.6	27	0.54	262	0.524
10	3	0.6	31	0.62	247	0.494

古典概型

定义：试验E中样本点是有限的，出现每一样本点的概率是相同。

$$P(A)=\frac{A\text{所包含的样本点数}}{S\text{中的样本点数}}$$

一袋中有8个球，编号为1 - 8，其中1 - 3号为红球，4 - 8号为黄球，设摸到每一球的可能性相等，从中随机摸一球，记A={ 摸到红球 }，求$P(A)$。

$S=\{1,2,\ldots,8\}$

$$A=\{1,2,3\}\Rightarrow P(A)=\frac{3}{8}$$

条件概率

3张奖券中只有1张能中奖，现分别由3名同学无放回地抽取，问最后一名同学抽到中奖奖券的概率是否比其他同学小?

Y表示抽到了，N表示木有抽中，所有的可能情况为: $\Omega=\{YNN, NYN, NNY\}$
B表示最后那个同学中了: $B=\{NNY\}$

有古典概型可知: $P(B)=\dfrac{n(B)}{n(\Omega)}=\dfrac{1}{3}$

一般用Ω表示所有基事件的集合

条件概率

- 如果已经知道第一个同学没抽中，那最后一名抽中的可能性会变吗?
- 第一名没中则: $A=\{NYN,NNY\}$
- B事件依旧表示最后那同学中了: $B=\{NNY\}$
- 那第一未中，第三中的事件发生的概率: $P(B\mid A)=\dfrac{n(B)}{n(A)}=\dfrac{1}{2}$

概率论

条件概率

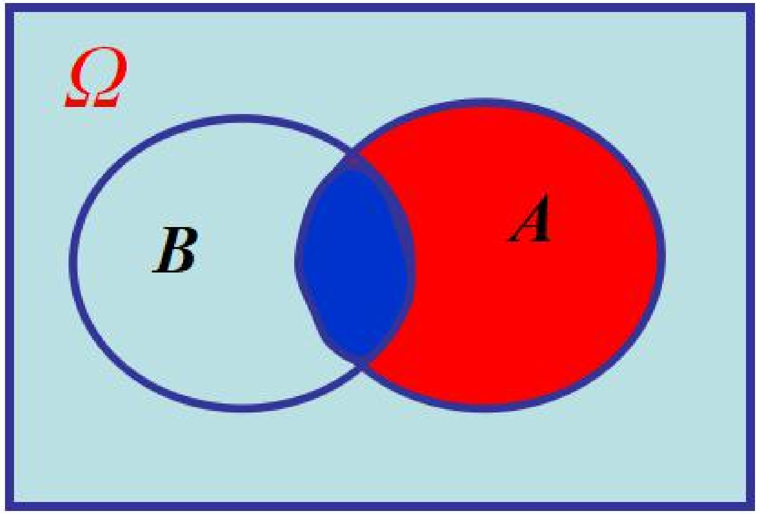
- 为什么结果不一样呢？什么变了？
- 未知第一个同学时，样本空间为：$\Omega=\{YNN,NYN,NNY\}$
- 知道第一同学未中时，样本空间为：$A=\{NYN,NNY\}$
- 但是第三个同学中奖的情况依旧只有一种：$\{NNY\}$

概率论

- 样本空间是什么样?

  - P(B)以试验下为条件,样本空间是 Ω

  - P(B|A)以A发生为条件,样本空间缩小为A



P(B |A)相当于把 A 看作新的样本空间求 A B 发生的概率

概率论

条件概率

P(B|A)的求解思路：$P(B|A)=\dfrac{n(AB)}{n(A)}$

因为已经知道事件A必然发生，所以只需在A发生的范围内考虑问题，即现在的样本空间为A。

因为在事件A发生的情况下事件B发生，等价于事件A和事件B同时发生，即AB发生。

$$P(B|A)=\frac{n(AB)/n(\Omega)}{n(A)/n(\Omega)}=\frac{P(AB)}{P(A)}$$

P(B|A)与P(AB)

- 相同点：事件A，B都发生了
- 不同点：样本空间不同：在P(B|A)中，事件A成为样本空间；在P(AB)中，样本空间仍为Ω。

例题

甲乙两地都位于长江下游，根据一百多年的气象记录，知道甲乙两地一年中雨天所占的比例分别为20%和18%，两地同时下雨的比例为12%，问：

（1）乙地为雨天时甲地也为雨天的概率是多少？

（2）甲地为雨天时乙地也为雨天的概率是多少？

设A={甲为雨天}， B={乙为雨天}则P(A)=20%，P(B)=18%，P(AB)=12%

（1）乙地为雨天时甲地也为雨天的概率是

$$P(A|B)=\frac{P(AB)}{P(B)}=\frac{12\%}{18\%}=\frac{2}{3}$$

（2）甲地为雨天时乙地也为雨天的概率是

$$P(B|A)=\frac{P(AB)}{P(A)}=\frac{12\%}{20\%}=\frac{3}{5}$$

例题

某厂生产的产品能直接出厂的概率为70%，余下的30%的产品要调试后再定，已知调试后有80%的产品可以出厂，20%的产品要报废。求该厂产品的报废率。

设 A={生产的产品要报废}

B={生产的产品要调试}

已知$P(B)=0.3$，$P(A|B)=0.2$，$P(A|\bar{B})=0$

$$P(A)=P(AB\cup A\bar{B})=P(AB)+P(A\bar{B})$$

$$=P(B)\cdot P(A|B)+P(\bar{B})\cdot P(A|\bar{B})$$

$$=0.3\times 0.2+0.7\times 0=6\%$$

独立性

- 设A，B为两随机事件，

  若$P(B|A)=P(B)$，即$P(AB)=P(A)\times P(B)$

  即$P(A|B)=P(A)$时，称A，B相互独立。

- 设$A_1, A_2, \cdots, A_n$为n个随机事件，若对$2 \le k \le n$,

  均有：$P\left(A_{i_1} A_{i_2} \cdots A_{i_k}\right)=\prod_{j=1}^{k} P\left(A_{i_j}\right)$

  则称$A_1, A_2, \cdots, A_n$相互独立

- 但是两两独立并不能得出相互独立！

例题

甲、乙两人同时向一目标射击，甲击中率为0.8，乙击中率为0.7，求目标被击中的概率。

设 A={甲击中},B={乙击中},C={目标被击中}

则：$C = A \cup B$，$P(C) = P(A) + P(B) - P(AB)$

∵ 甲、乙同时射击，其结果互不影响，

∴ A，B相互独立

$$\Rightarrow P(C) = 0.7 + 0.8 - 0.56 = 0.94$$

独立试验

重复独立试验：在相同的条件下，将试验E重复进行，且每次试验是独立进行的，即每次试验各种结果出现的概率不受其他各次试验结果的影响。

n重伯努利试验：若一试验的结果只有两个A和$\bar{A}$, 在相同的条件下, 将试验独立地重复进行n次, 则称这n次试验所组成的试验为n重复伯努利试验或伯努利概型。

将一枚均匀的骰子连续抛掷3次，考察六点出现的次数及相应的概率。

设六点出现的次数为X, 设第i次抛掷中出现点6的事件为

$A_k, k=1,2,3,$ 则

$$P(X=0)=P(\overline{A}_1\overline{A}_2\overline{A}_3)=(5/6)^3=0.578704=C_3^0(1/6)^0(5/6)^3$$

$$P(X=1)=P(A_1\overline{A}_2\overline{A}_3\cup\overline{A}_1A_2\overline{A}_3\cup\overline{A}_1\overline{A}_2A_3)$$

$$=3(1/6)(5/6)^2=0.347222=C_3^1(1/6)(5/6)^2$$

$$P(X=2)=P(A_1A_2\overline{A}_3\cup A_1\overline{A}_2A_3\cup\overline{A}_1A_2A_3)$$

$$=3(1/6)^2(5/6)=0.069444=C_3^2(1/6)^2(5/6)$$

$$P(X=3)=P(A_1A_2A_3)$$

$$=(1/6)^3=0.004630=C_3^3(1/6)^3(5/6)^0$$

因此,我们有

$$P(X=k)=C_3^k(1/6)^k(5/6)^{3-k},\quad k=0,1,2,3$$

n重伯努利试验

如果每次试验中事件A发生的概率为，

$$p(0 < p < 1)$$

则在n次贝努里试验中事件A恰好发生k次的概率为

$$P_n(k) = C_n^k p^k q^{n-k}, k = 0,1,\cdots,n$$

其中 $q = 1 - p$

二维随机变量

以前我们只关心一个指标，现在要更操心了，例如根据学生的身高（X）和体重（Y）来观察学生的身体状况。

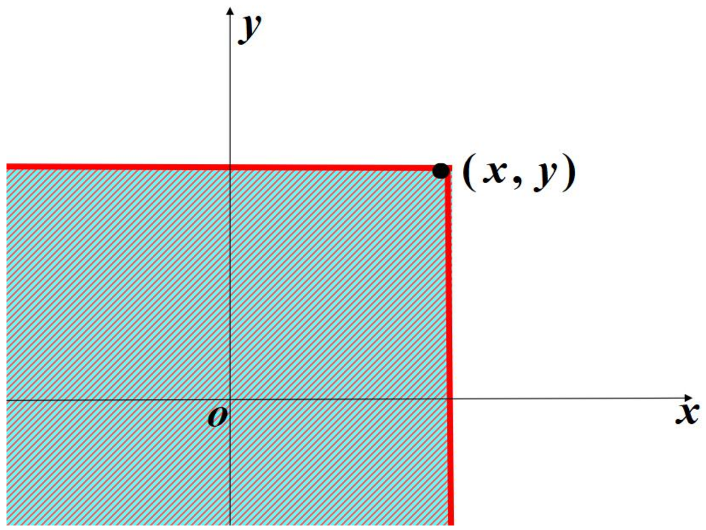
这就不仅仅是X和Y各自的情况，还需要了解其相互的关系。

二维随机变量

- 二维随机变量的联合函数：若（X,Y）是随机变量，对于任意的实数x,y

$$F(x,y)=P\{(X\le x)\cap(Y\le y)\}$$

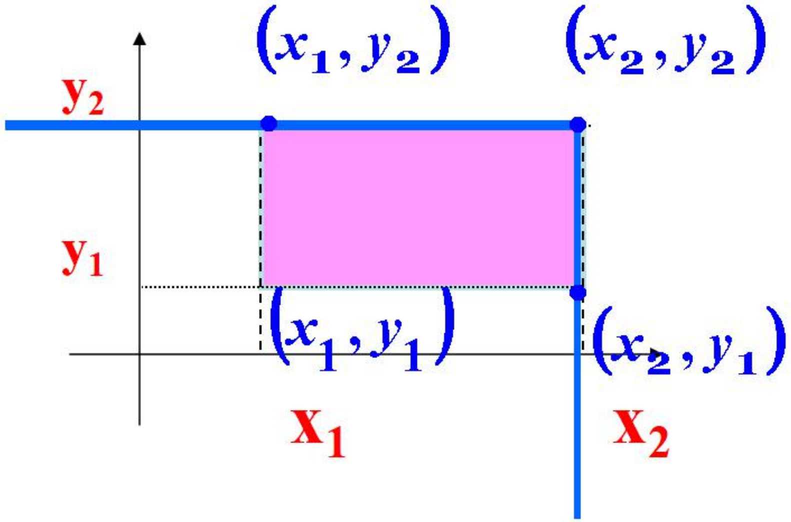
- F(x,y)表示随机点（X,Y）在以（x,y）为顶点且位于该点左下方无穷矩形内的概率。



二维随机变量

用联合分布函数F(x,y)表示矩形域概率

$$P\left(x_1 < X \leq x_2,\ y_1 < Y \leq y_2\right)$$
$$= F(x_2,y_2) - F(x_2,y_1) - F(x_1,y_2) + F(x_1,y_1)$$



$F(x_2,y_2)$

二维随机变量

性质（1）	$F(x,y)$分别关于X和Y__单调不减；____.
性质（2）	$\underline{\quad 0 \quad} \le F(x,y) \le \underline{\quad 1 \quad}$. $F(x,-\infty)=\underline{\quad 0 \quad}$；$F(-\infty,y)=\underline{\quad 0 \quad}$. $F(-\infty,-\infty)=\underline{\ 0\ }$；$F(+\infty,+\infty)=\underline{\ 1\ }$.
性质（3）	$F(x,y)$分别关于X和Y__右连续；____.
性质（4）	$\forall\ (x_2,y_2)$，(x_1,y_1)，$x_1<x_2$，$y_1<y_2$，有 $F(x_2,y_2)-F(x_2,y_1)-F(x_1,y_2)+F(x_1,y_1)\ge 0$

二维随机变量的概率分布

若二维随机变量(X,Y)全部可能取到的不同值是有限对或可列无限对，则称(X,Y)是离散型随机变量。

离散型随机变量的联合概率分布：

设(X,Y)所有可能取值为

(x_i, y_i)，$i, j = 1, 2, \cdots$

称$P\{X = x_i, Y = y_j\} = P(x_i, y_j) = p_{ij}$，$i, j = 1, 2, \cdots$

$X \backslash Y$	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}	$\cdots$
$\vdots$	$\cdots$		$\cdots$		$\cdots$
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$
$\vdots$	$\cdots$		$\cdots$		$\cdots$

$p_{ij} \geq 0$，$i, j = 1, 2, \cdots$

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} P_{ij} = 1$$

为二维离散型随机变量(X,Y)的联合概率分布。

设随机变量X在1、2、3、4四个整数中等可能地取 一个值，另一个随机变量Y在1~X中等可能地取一整数值，试求(X, Y)的联合概率分布。

$(X=i, Y=j)$的取值情况为：i=1,2,3,4；j取不大于i的正整数。

$$P\{X=i, Y=j\}=P\{X=i\}P\{Y=j \mid X=i\}=\frac{1}{4}\cdot\frac{1}{i}$$

$i=1,2,3,4;\ j \le i$

即(X, Y)的联合概率分布为：

$Y \backslash X$	1	2	3	4
1	1/4	1/8	1/12	1/16
2	0	1/8	1/12	1/16
3	0	0	1/12	1/16
4	0	0	0	1/16

概率论

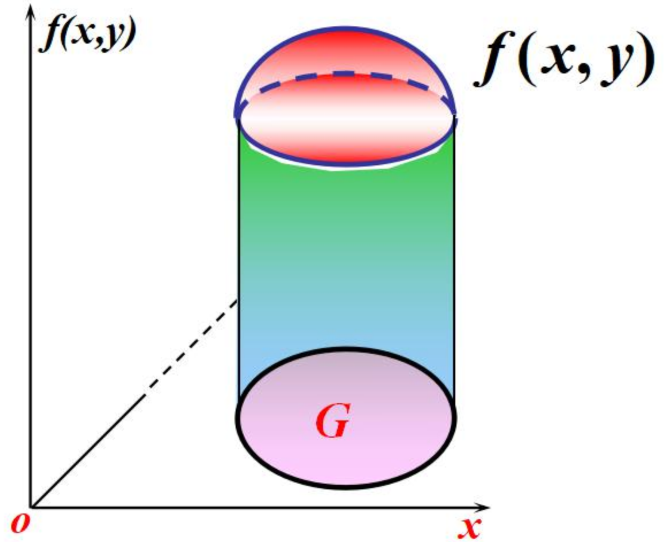
二维连续型随机变量

二维随机变量(X,Y)的分布函数$F(x,y)$如果存在非负函数$f(x,y)$

对于任意x,y有：$F(x,y)=\int_{-\infty}^{y}\int_{-\infty}^{x}f(u,v)dudv$

称(X,Y)为连续型的二维随机变量，$f(x,y)$ 为其概率密度。

$$P\{(X,Y)\in G\}=\iint_G f(x,y)dxdy$$



概率论

设二维随机变量(X,Y)具有概率密度：

$$f(x,y)=\begin{cases}ke^{-(2x+3y)}, & x>0,\ y>0\\ 0, & \text{其它}\end{cases}$$

(1) 求 常 数 k；

(2) 求分布函数$F(x,y)$;

(3) 求$P(Y\le X)$的概率

(1)利用$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)dxdy=1$,得

$$k\int_0^{\infty}e^{-2x}dx\int_0^{\infty}e^{-3y}dy=k/6=1$$

$$\Rightarrow k=6$$

$$f(x,y)=\begin{cases}6e^{-(2x+3y)}, & x>0,\ y>0\\ 0, & \text{其它}\end{cases}$$

$$(2)\ F(x,y)=\int_{-\infty}^{y}\int_{-\infty}^{x}f(u,v)dudv=\begin{cases}\int_0^y\int_0^x 6e^{-(2u+3v)}dudv, & x>0,y>0\\ 0, & \text{其它}\end{cases}$$

$$=\begin{cases}\int_0^x 2e^{-2u}du\int_0^y 3e^{-3v}dv, & x>0,y>0\\ 0, & \text{其它}\end{cases}=\begin{cases}(1-e^{-2x})(1-e^{-3y}), & x>0,y>0\\ 0, & \text{其它}\end{cases}$$

$$(3)\ P(Y\le X)=\int_0^{\infty}\int_y^{\infty}6e^{-(2x+3y)}dxdy=\int_0^{\infty}3e^{-3y}(-e^{-2x}|_y^{\infty})dy$$

$$=\int_0^{\infty}3e^{-3y}e^{-2y}dy=\int_0^{\infty}3e^{-5y}dy=-\frac{3}{5}e^{-5y}|_0^{\infty}=\frac{3}{5}$$

边缘分布

- 边缘分布函数：二维随机变量(X,Y)作为整体，有分布函数$F(x,y)$,其中，X和Y都是随机变量，它们的分布函数记为：$F_X(x),F_Y(y)$ 称为边缘分布函数。

- 在分布函数$F(x,y)$中令$y\to+\infty$，就能得到$F_X(x)$

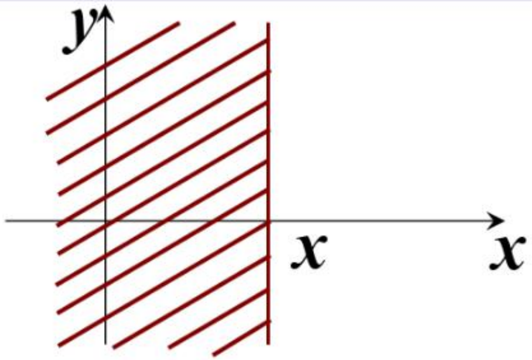
$$F_X(x)=P\{X\le x\}=P\{X\le x,Y<+\infty\}=F(x,+\infty)$$

同理得：$F_Y(y)=P\{Y\le y\}=F(+\infty,y)$

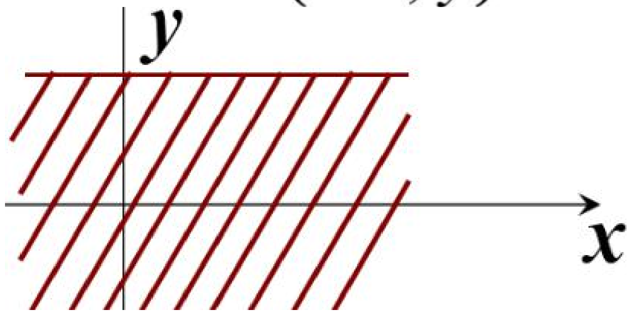
概边缘分布

由联合分布函数可以得到边缘分布函数：

$$F_X(x) = P(X \le x)$$
$$= P(X \le x, Y < +\infty)$$
$$= F(x, +\infty)$$



$$F_Y(y) = P(Y \le y)$$
$$= P(X < +\infty, Y \le y)$$
$$= F(+\infty, y)$$



离散型的边缘分布

对于离散型随机变量(X,Y)，分布律为

$$P\{X=x_i,\ Y=y_j\}=p_{ij},\ i,j=1,2,\cdots$$

X,Y的边缘分布律为:

$$P\{Y=y_j\}=P\{X<+\infty,\ Y=y_j\}=\sum_{i=1}^{\infty}p_{ij}\overset{\text{记为}}{=\!=}p_{\bullet j}\quad j=1,2,\cdots$$

$$P\{X=x_i\}=P\{X=x_i,\ Y<+\infty\}=\sum_{j=1}^{\infty}p_{ij}\overset{\text{记为}}{=\!=}p_{i\bullet}\quad i=1,2,\cdots$$

$X \backslash Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	$P(X=x_i)$
x_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$	$p_{1\cdot}$
x_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$	$p_{2\cdot}$
$\vdots$	$\cdots$		$\cdots$		$\cdots$	$\vdots$
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$	$p_{i\cdot}$
$\vdots$	$\cdots$		$\cdots$		$\cdots$	$\vdots$
$P(Y=y_j)$	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot j}$	$\cdots$	1

连续型的边缘概率密度

对于连续型随机变量(X,Y)，概率密度为 $f(x,y)$

X,Y的边缘概率密度为：$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy$

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx$$

事实上，$F_X(x)=F(x,+\infty)=\int_{-\infty}^{x}\left[\int_{-\infty}^{+\infty}f(t,y)dy\right]dt$

$$=\int_{-\infty}^{x}f_X(t)dt$$

同理：

$$F_Y(y)=F(+\infty,y)=\int_{-\infty}^{y}\left[\int_{-\infty}^{+\infty}f(x,t)dx\right]dt$$

$$=\int_{-\infty}^{y}f_Y(t)dt$$

对一群体的吸烟及健康状况进行调查，引入随机变量

X和Y如下：$X=\begin{cases}0, & \text{健康}\\ 1, & \text{一般}\\ 2, & \text{不健康}\end{cases}$，$Y=\begin{cases}0, & \text{不吸烟}\\ 10, & \text{一天吸烟不多于 15支}\\ 20, & \text{一天吸烟多于15支}\end{cases}$

根据调查结果,得 (X,Y)的如下的联合概率分布：

$X \backslash Y$	0	10	20
0	0.35	0.04	0.025
1	0.025	0.15	0.04
2	0.020	0.10	0.25

(1) 试写出关于X和Y的边缘概率分布；

(2) 求$P\{X=2\mid Y=20\}$的值。

(1)由题意可得：

X	0	1	2
p	0.415	0.215	0.370

Y	0	10	20
p	0.395	0.290	0.315

(2)$P\{X=2\mid Y=20\}=\dfrac{0.25}{0.315}=0.794$

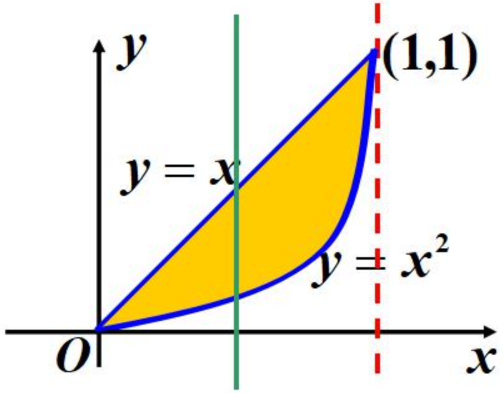
设随机变量 X 和 Y 具有联合概率密度

$$f(x,y)=\begin{cases}6, & x^2 \le y \le x,\\ 0, & \text{其他}.\end{cases}$$

求边缘概率密度 $f_X(x), f_Y(y)$.

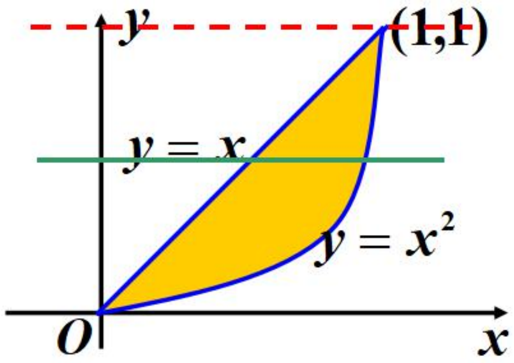
$$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}y$$

$$=\begin{cases}\int_{x^2}^{x} 6dy & 0 \le x \le 1\\ 0 & \text{其他}\end{cases}=\begin{cases}6(x-x^2), & 0 \le x \le 1,\\ 0, & \text{其他}.\end{cases}$$



$$f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}x$$

$$=\begin{cases}\int_{y}^{\sqrt{y}} 6dx & 0 \le y \le 1\\ 0 & \text{其他}\end{cases}=\begin{cases}6(\sqrt{y}-y), & 0 \le y \le 1,\\ 0, & \text{其他}.\end{cases}$$



期望

- 很多时候我们需要掌握的是一个具体的指标，比如粮食的平均产量，一个地区家庭的平均年收入，贫富差距等。

- 离散型随机变量X的分布律为：$P(X=x_k)=p_k \quad k=1,2,\cdots$

  若级数 $\sum_{k=1}^{\infty} x_k p_k$ 绝对收敛，则称其为随机变量X的数学期望，$E(X)=\sum_{k=1}^{\infty} x_k p_k$

- 投骰子的期望等于多少？

期望

连续型随机变量X的概率密度为$f(x)$，若积分$\int_{-\infty}^{+\infty} xf(x)dx$绝对收敛

则称积分的值$\int_{-\infty}^{+\infty} xf(x)dx$为随机变量X的数学期望。$E(X)=\int_{-\infty}^{+\infty} xf(x)dx$

随机变量X满足于均匀分布，求其期望。

$$X \sim f(x)=\begin{cases} \dfrac{1}{b-a}, & a<x<b, \\ 0, & \text{其它}, \end{cases} \longrightarrow E(X)=\frac{a+b}{2}$$

$$E(X)=\int_a^b \frac{x}{b-a}dx=\frac{a+b}{2};$$

二维情况

离散型：若 $(X, Y) \sim P\{X=x_i, Y=y_j\}=p_{ij},\ i, j=1, 2, \ldots$，则$Z=g(X, Y)$的期望

$$E(Z)=E[g(X,Y)]=\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}g(x_i,y_j)p_{ij}$$

连续型：若二维连续型随机变量(X,Y)的概率密度为：$z=g(x,y)$

设$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f(x,y)dxdy$绝对收敛,则有

$$E(Z)=E(g(X,Y))=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f(x,y)dxdy$$

设二维随机变量 (X,Y) 的联合分布律为

$X \backslash Y$	0	1	2
0	0.1	0.25	0.15
1	0.15	0.2	0.15

求随机变量 $Z=\sin\frac{\pi(X+Y)}{2}$ 的数学期望。

$$E(Z)=E[\sin\frac{\pi(X+Y)}{2}]=\sin\frac{\pi(0+0)}{2}\times 0.1+\sin\frac{\pi(1+0)}{2}\times 0.15$$

$$+\sin\frac{\pi(0+1)}{2}\times 0.25+\sin\frac{\pi(1+1)}{2}\times 0.2+\sin\frac{\pi(0+2)}{2}\times 0.15$$

$$+\sin\frac{\pi(1+2)}{2}\times 0.15=0.25$$

期望

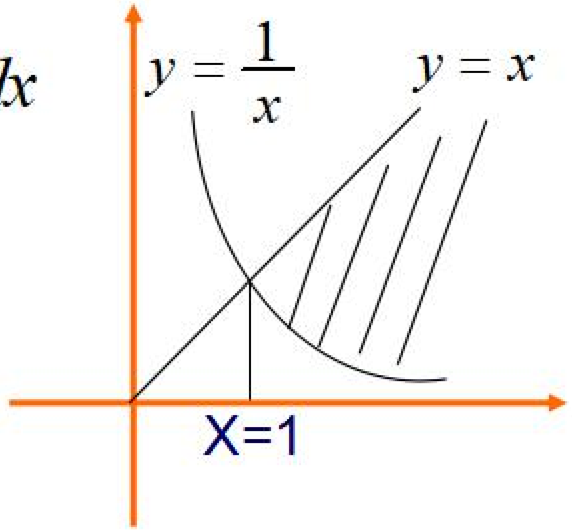
随机变量（X,Y）的概率密度为：$f(x,y)=\begin{cases}\dfrac{3}{2x^3y^2} & \dfrac{1}{x}<y<x, x>1\\ 0 & \text{其它}\end{cases}$

求数学期望$E(Y)$

$$E(Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yf(x,y)dydx=\int_{1}^{+\infty}\int_{\frac{1}{x}}^{x}\frac{3}{2x^3y}dydx$$

$$=\frac{3}{2}\int_{1}^{+\infty}\frac{1}{x^3}\ln y\Big|_{\frac{1}{x}}^{x}dx=3\int_{1}^{+\infty}\frac{\ln x}{x^3}dx$$

$$=-\frac{3}{2}\frac{\ln x}{x^2}\Big|_{1}^{+\infty}+\frac{3}{2}\int_{1}^{+\infty}\frac{1}{x^3}dx=\frac{3}{4}$$



数学期望的性质

1. 设C是常数，则有$E(C)=C$

2. 设X是一个随机变量， C是常数，则有 $E(CX)=CE(X)$

3. 设X,Y是两个随机变量，则有$E(X+Y)=E(X)+E(Y)$

   将上面三项合起来就是： $E(aX+bY+c)=aE(X)+bE(Y)+c$

4. 设X,Y是相互独立的随机变量，则有 $E(XY)=E(X)E(Y)$

期望

一民航送客车载有20位旅客自机场出发，旅客有10个 车站可以下车，如到达一个车站没有旅客下车就不停车，以X表示停车的次数,求E(X)。

$$X_i = \begin{cases} 0 & \text{第}i\text{站没有人下车} \\ 1 & \text{第}i\text{站有人下车} \end{cases} \quad i = 1, 2, \cdots, 10$$

$$X = X_1 + X_2 + \cdots + X_{10}$$

$$E(X_i) = P(X_i = 1) \quad = P(\text{第}i\text{站有人下车}) \quad = 1 - \left(\frac{9}{10}\right)^{20}$$

$$E(X) = E(X_1) + E(X_2) + \cdots + E(X_{10})$$

$$= 10\left[1 - \left(\frac{9}{10}\right)^{20}\right] = 8.784(\text{次})$$

方差

- 数学期望反映了随机变量的取值水平，衡量随机变量相对于数学期望的分散程度则的另一个数字特征。

- X为随机变量，如果 $E[X-E(X)]^2$ 存在，则称其为X的方差，记作D(X)

$$D(X)=E[X-E(X)]^2=E\left(X^2\right)-[E(X)]^2$$

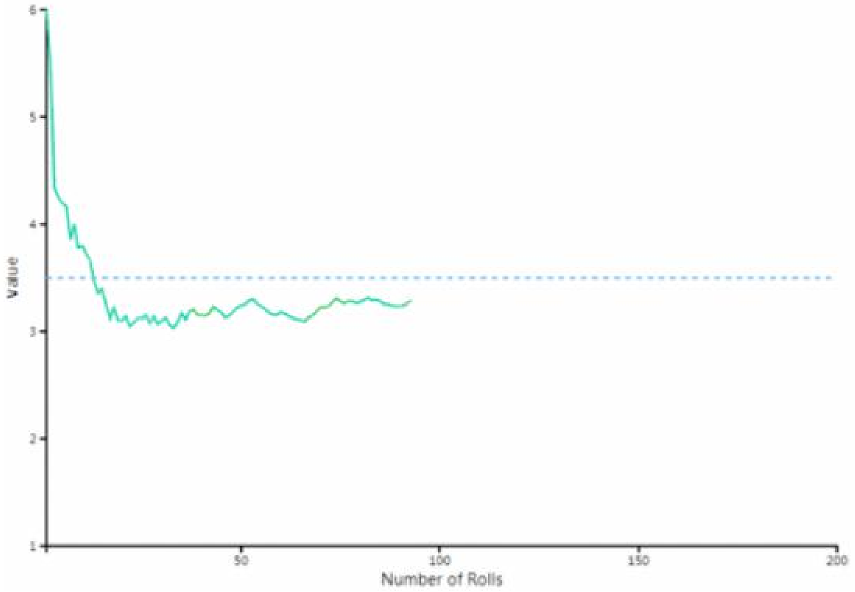
概率论

大数定理：

在试验不变的条件下，重复试验多次，随机事件的频率近似于它的概率。

小的样本试验不足以以偏概全因为有一些局限。

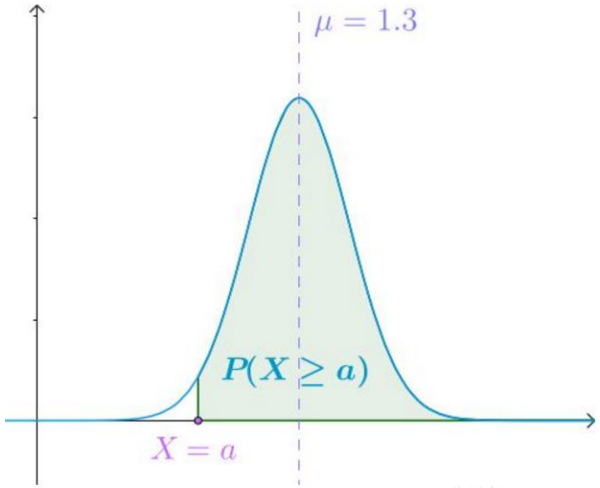
当我们投掷骰子的时，期望会等于多少呢？



概率论

马尔科夫不等式

$P(X \geq a) \leq \dfrac{E(X)}{a}$ $X \geq 0, a > 0$ 怎么得到的呢？



$$\left.\begin{matrix} X \geq 0 \\ X \geq a \end{matrix}\right\} \Longrightarrow \frac{X}{a} \geq 1$$

如何扩大这个面积呢？ $P(X \geq a) = \displaystyle\int_a^{+\infty} f(x)dx \leq \int_a^{+\infty} \frac{X}{a} f(x)dx$

根据期望有：$E(\dfrac{X}{a}) = \displaystyle\int_{-\infty}^{a} \frac{X}{a} f(x)dx + \int_a^{+\infty} \frac{X}{a} f(x)dx$ 由于：$\displaystyle\int_{-\infty}^{a} \frac{X}{a} f(x)dx \geq 0$

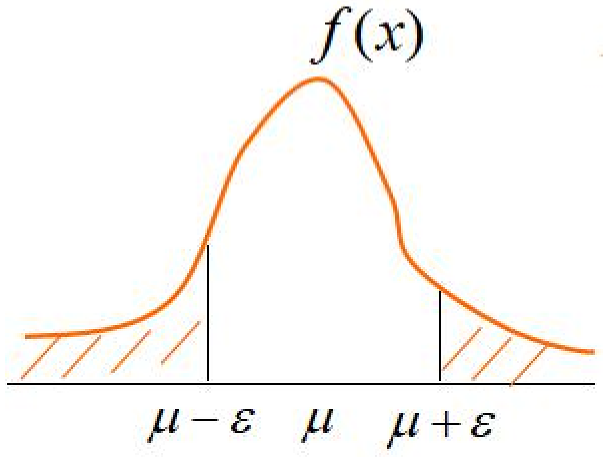
$P(X \geq a) \leq \displaystyle\int_a^{+\infty} \frac{X}{a} f(x)dx \leq E(\frac{X}{a})$ 即：$P(X \geq a) \leq E\left(\dfrac{X}{a}\right) = \dfrac{E(X)}{a}$

选自马同学解释

概率论

切比雪夫不等式

$$P\{|X-E(X)|\geq\varepsilon\}\leq\frac{\sigma^2}{\varepsilon^2}\quad P\{|X-E(X)|<\varepsilon\}\geq 1-\frac{\sigma^2}{\varepsilon^2}$$



将 $|X-\mu|$ 带入马尔科夫不等式 $P(X\geq a)\leq\frac{E(X)}{a}$

可得: $P(|X-\mu|>a)\leq\frac{E(|X-\mu|)}{a}$ 即: $P((X-\mu)^2\geq a^2)\leq\frac{E((X-\mu)^2)}{a^2}=\frac{\sigma^2}{a^2}$

切比雪夫不等式

在n重贝努里试验中，若已知每次试验事件A出现的概率为0.75，试利用契比雪夫不等式估计n，使A出现的频率在0.74至0.76之间的概率不小于0.90。

设在n重贝努里试验中，事件A出现的次数为X，

则$X \sim b(n, 0.75)$，

$E(X) = np = 0.75n, D(X) = npq = 0.1875n,$

又 $f_n(A) = \dfrac{X}{n}$ 而$P\left\{0.74 < \dfrac{X}{n} < 0.76\right\} = P\{|X - 0.75n| < 0.01n\} \geq 1 - \dfrac{0.1875n}{(0.01n)^2} = 1 - \dfrac{1875}{n} \geq 0.90$

$\Rightarrow n \geq 18750$

概率论

中心极限定理

- 样本的平均值约等于总体的平均值。不管总体是什么分布，任意一个总体的样本平均值都会围绕在总体的整体平均值周围，并且呈正态分布。

- 描述的是一个实际的现象，有了这个定理就能解决很多问题了，比如我们可以通过对样本进行观察，得出总体的情况。

http://onlinestatbook.com/stat_sim/sampling_dist/index.html