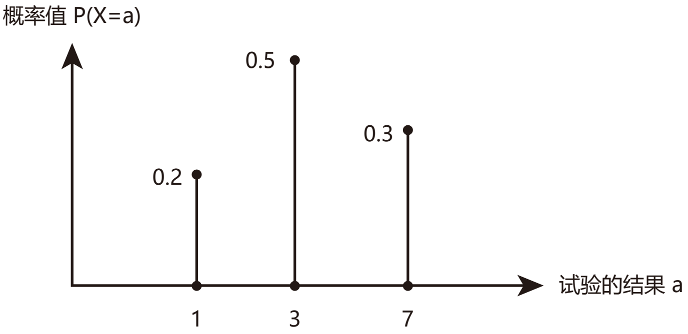
目录

文件名

1 离散型数据 - 概率函数 (概率质量函数)：Probability mass function (PMF)

描述"离散型数据"的概率分布情况的曲线, 称为"概率质量函数"（PMF）.



概率函数曲线 PMF

比如，掷骰子, 不同点朝上的概率为: $\underbrace{p_i}_{\text{骰子在第}i\text{点的概率}} = \underbrace{P(X=a_i)}_{\text{随机变量}X\text{在第}i\text{点时的概率}}, \quad i \in 1,2,3,4,5,6$

在上面这个函数里：
- 自变量X：是"随机变量"的取值.
- 因变量 p_i 是"自变量X所取到某个值 a_i"的概率.

从公式上来看, **"概率函数", 一次只能表示一个取值的概率.** 比如 $P(X=1)=1/6$ 就表示: 当随机变量X 取值为 1时 (即骰子投到点数为1时) 的概率为1/6. 所以说, 它一次只能代表"随机变量的一个取值"的概率. **(即: P后面的小括号里, 只能写成 X= ..., 而不能写成 X¿... 或 X¡..., 大于小于符号这些, 就是属于"累加函数"了.)**

典型的"离散概率分布"包括: 伯努利分布，二项分布，几何分布，泊松分布等.

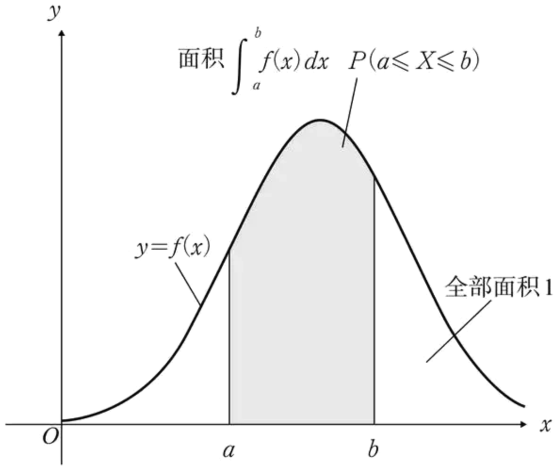
---

2 连续型数据 - 概率函数 (概率密度函数)：Probability Density function (PDF)

"连续型数据"的概率分布, 称为"概率密度函数"（PDF）.

"概率密度函数"的某区间上的概率值 = 该区间的函数曲线段, 与x坐标轴之间围成的面积. 实际上就是对"概率密度函数"进行定积分.

$\int_{-\infty}^{+\infty}$ 概率函数$f(x)dx = 1$



例

有概率函数 $f(x) = \begin{cases} kx+1 & (0 \le x \le 2) \\ 0 & (x\text{是其他的话}) \end{cases}$

(1)求k.
根据"概率密度函数 f(x)"的性质, 有 $\int_{-\infty}^{+\infty} f(x)dx = 1$.
既然本题中, x只在(0,2)区间上才有概率值(=kx+1); 而在其他定义域区间上, 概率值都是=0. 这就说明, 该概率函数的全部概率值("求和=1"), 就只在(0,2)区间上.

即:

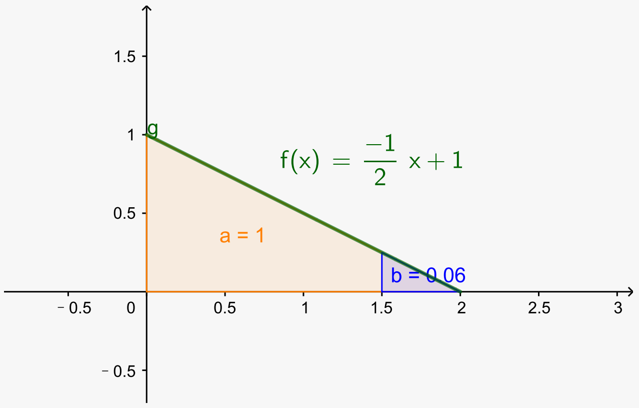
$$\int_0^2 (kx+1)dx = 1$$
$$\int_0^2 \left(\underbrace{k}_{\text{可视为常数}} x\right) dx + \int_0^2 1dx = 1$$
$$k\left(\frac{1}{1+1}x^{1+1}\,\Big|_0^2\right) + x\,\Big|_0^2 = 1$$
$$k\left(\frac{x^2}{2}\,\Big|_0^2\right) + 2 = 1$$
$$k\frac{2^2}{2} + 2 = 1$$
$$k = -\frac{1}{2}$$

(2)求

$$P\{x \le 2\}$$
$$= P\{-\infty < x < 2\}$$
$$= \int_{-\infty}^{2} \underbrace{f(x)}_{=kx+1,\text{而}k\text{上面已算出}=-\frac{1}{2}} dx$$
$$\leftarrow \text{因为}x\text{只在}\,(0,2)\,\text{区间上有值,所以本处的}\,(-\infty,2)\,\text{其实就可直接写成}\,(0,2)$$
$$= \int_{-\infty}^{2} \left(-\frac{1}{2}x+1\right)dx = 1$$

(3)求

$$P\{1.5<x<2.5\}$$
$$=\int_{1.5}^{2.5}\underbrace{f(x)}_{=kx+1,\text{而}k\text{上面已算出}=-\frac{1}{2}}dx$$
$\leftarrow$ 因为x只在 $(0,2)$ 区间上有值,所以本处的 $(1.5,2.5)$ 其实就可直接写成 $(1.5,2)$
$$=\int_{1.5}^{2}\left(-\frac{1}{2}x+1\right)dx=0.0625$$



2.1 连续型随机变量, 取任一"个别值"的概率, 都为0.

比如, 在一段区间上, 投掷质点(无面积, 0维), 该质点砸中任何一个数值的概率, 就是为0.
你可以倒过来想: 如果"该质点能砸中某个数值"的概率, 是可以给出的, 比如是 0.000001%, 那"一段区间"上是有无穷多的点的, 0.000001% 乘以无穷多, 一定是会超过 100%的, 这就违反了概率不能超过1 的定义. 所以, "质点投中任何位置处"的概率, 都无法给出, 是0.

所以, 概率为0 的事件, 未必是"不可能事件". (如, 扔质子)
概率为1 的事件, 未必是"必然事件".

2.2 如何求 PDF在某一x点处的y值(概率)?

那么对于"连续性概率函数", 要求它在"x=某一点处"的概率, 该怎么求呢? $\rightarrow$ 要用"极限"的概念来求.

即:

$$\lim_{\Delta x\to 0}\frac{P\{x<\text{随机变量}X<(x+\Delta x)\}}{\Delta x}$$
$$=\lim_{\Delta x\to 0}\frac{\int_x^{x+\Delta x}f(x)dx}{\Delta x}\leftarrow\text{分子分母是}\frac{0}{0}\text{型,用洛必达法则来做.}$$
$$\text{即:}P\{x<\text{随机变量}X<(x+\Delta x)\}\approx\underbrace{f(x)}_{y\text{高}}\underbrace{\Delta x}_{x\text{宽}}$$

所以, 千万不要误认为: "概率密度函数"在某一x点处的y值, 是该x点的概率. 事实上, **概率密度函数在某点的函数值, 是概率在该点的"变化率"(即斜率, 导数)!** 因为对于"连续性数据"来说, 任何一点的概率都是0. 所以任何一点是不存在概率的! 我们只能得到"概率密度函数"在该x点处的斜率. 这个就类似于导数的概念. 导数代表着"该点处切线的斜率"!

典型的"连续概率分布"包括: 正态分布, 指数分布等.