目录

文件名

第一部分　随机变量 random variable

随机变量: 常用大写字母X,Y,Z 或希腊字母来表示.
随机变量的取值：用小写字母 x,y,z等表示.

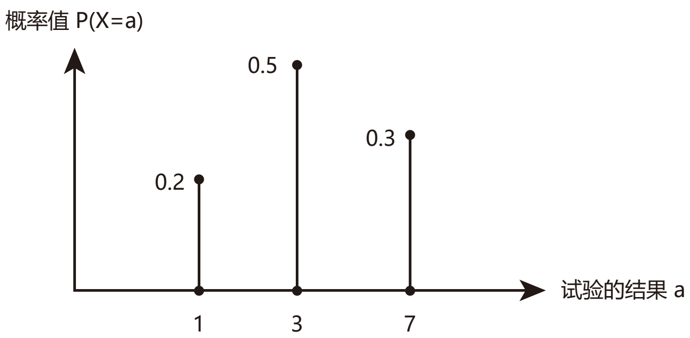
比如, 随机变量X, 其=a的话, 我们就把这个事件记作 $\{X = a\}$. 其概率就是 $P\{X = a\}$, 或写作 $P(X = a)$.

离散型随机变量	其值是 "有限可列举"的, 如, 为n个有限取值：$\{x_1, x_2, ..., x_n\}$).
连续型随机变量	其值无法逐个列举(即是无穷无尽个的), 是一段区间. 只能写成: $X = \{x \mid a \le x \le b\}, -\infty < a < b < \infty$.

第二部分　概率函数：PMF (离散型数据的, 称为"概率质量函数 Probability mass function") & PDF (连续性数据的, 称为"概率密度函数 Probability Density Function")

1　离散型数据的 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: $\underbrace{p_i}_{\text{骰子在第}i\text{点的概率}} = \underbrace{P(X = a_i)}_{\text{随机变量}X\text{在第}i\text{点时的概率}}$, $i \in 1, 2, 3, 4, 5, 6$

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.

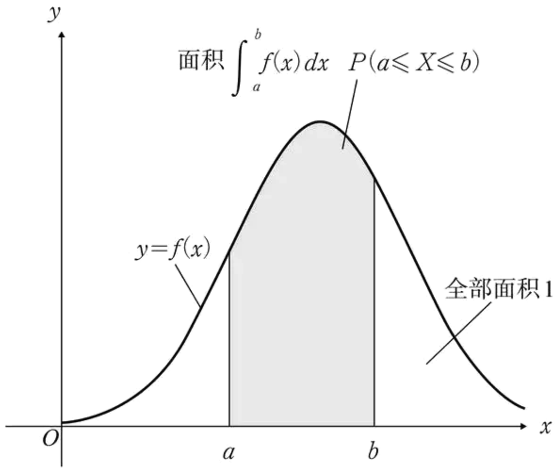
---

2 连续性数据的 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:

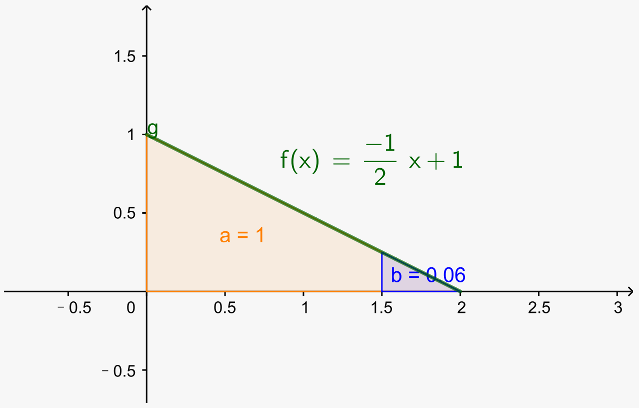
$$\begin{aligned}
&\int_0^2 (kx+1)dx = 1\\
&\int_0^2 \left(\underbrace{k}_{\text{可视为常数}} x\right)dx + \int_0^2 1dx = 1\\
&k\left(\frac{1}{1+1}x^{1+1}\mid_0^2\right) + x\mid_0^2 = 1\\
&k\left(\frac{x^2}{2}\mid_0^2\right) + 2 = 1\\
&k\frac{2^2}{2} + 2 = 1\\
&k = -\frac{1}{2}
\end{aligned}$$

(2)求

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

(3)求

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



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:

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

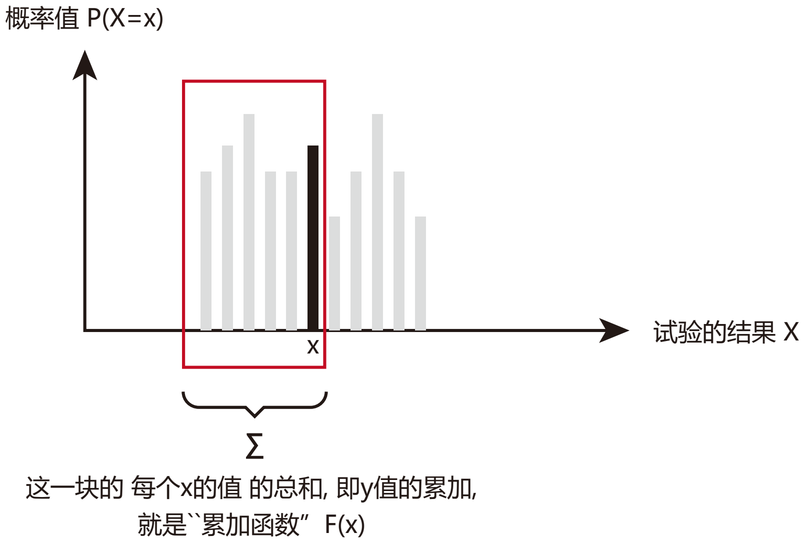
所以, 千万不要误认为: "概率密度函数"在某一x点处的y值, 是该x点的概率. 事实上, **概率密度函数在某点的函数值, 是概率在该点的"变化率"(即斜率, 导数)!** 因为对于"连续性数据"来说, 任何一点的概率都是0. 所以任何一点是不存在概率的! 我们只能得到"概率密度函数"在该x点处的斜率. 这个就类似于导数的概念. 导数代表着"该点处切线的斜率"!

典型的"连续概率分布"包括: 正态分布, 指数分布等.

第三部分 累积函数 $F\left(x\right)=P\left\{X\leq x\right\}=\sum_{x_k\leq x}p_k$: Cumulative Distribution Function (CDF) $\leftarrow$ 是对"概率函数"值的累加结果

对于随机变量, 我们通常关心的, 并不是它取某个值的概率(即我们并不关心它的分布律), 而是更关心它落在某个区间内的概率.
比如, 对某考试, 我们更关心的是"不及格的总人数", 和比如 "分数⩾80分的总人数".

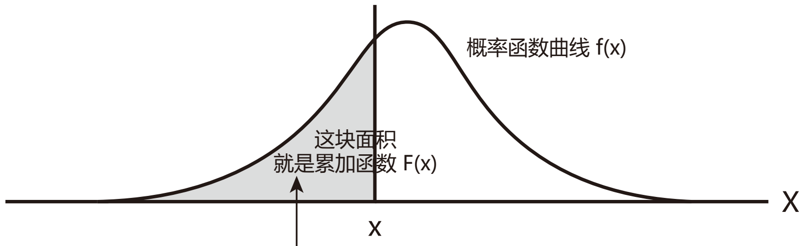
累积函数 Cumulative Distribution Function (CDF) $\leftarrow$ 是对"概率函数"值的累加结果 . 即对"概率密度函数"的积分.



在这些个区间段所占的概率值, 就是用“累加函数”(又叫“分布函数”)来表示的. 即:
(随机变量$X \leq$ 自变量x) $= \underbrace{F(x)}_{\text{累加函数}}$ $\leftarrow$ 它表示随机变量X 落在 (-∞, x] 这段区间上的概率.

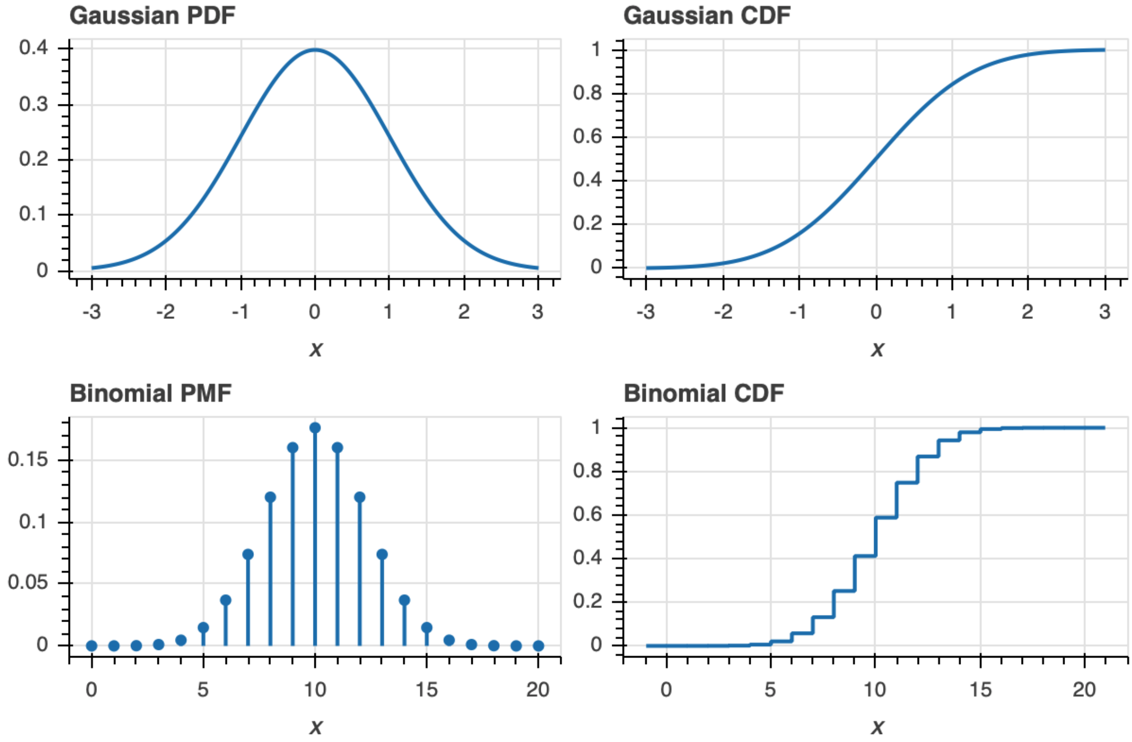
$$\boxed{\text{累加函数} F(x) = P\{X \leq x\} = \sum_{x_k \leq x} p_k}$$

累加函数 F(x) 就是 “X取≤x 的所有值x_k” 的概率之和.
P(X≤x) 即“X的取值不超过x” 的概率. 这里P后面写()或{}都行, 意思是一样的.



$$F(x) = P(X \leq x) = S_{\text{阴影部分面积}} = \int_{-\infty}^{x} f(t)d$$

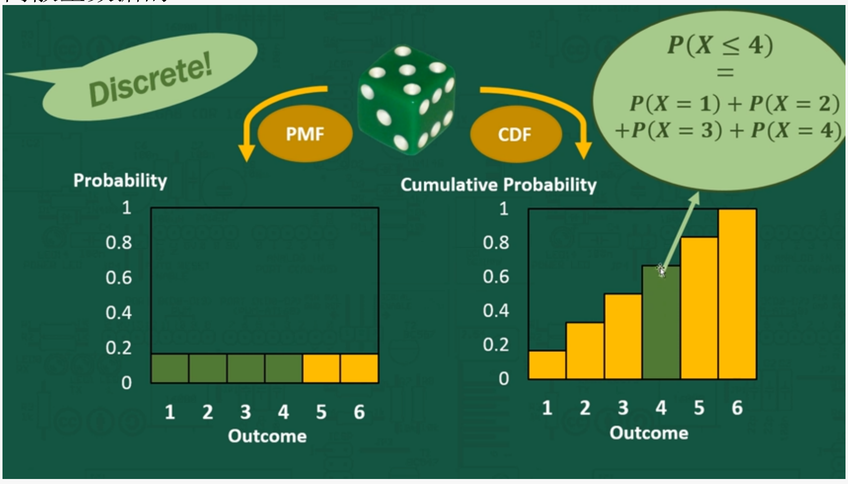
下图, 左边两张是“概率函数”, 右边两张就是“累加函数 CDF”.



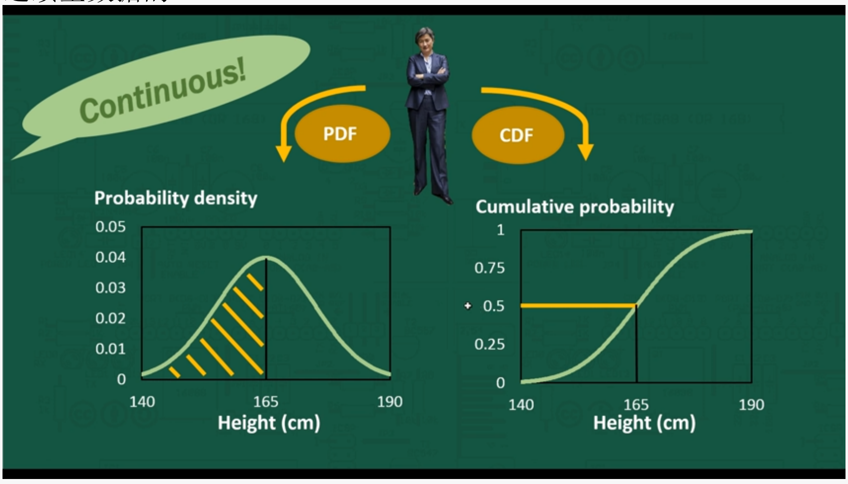
例

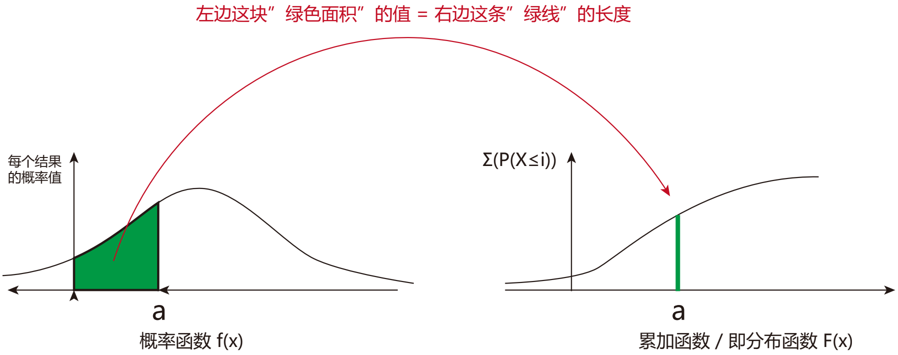
下面的图，左边是“概率函数”，右边是“累加函数”.

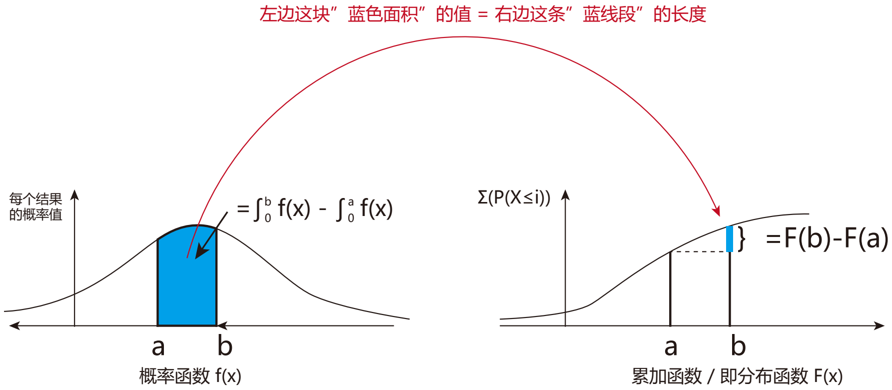
离散型数据的:

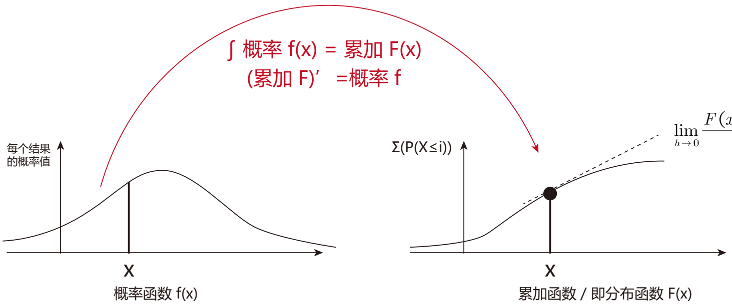


连续型数据的:





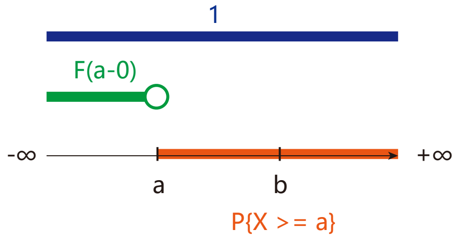
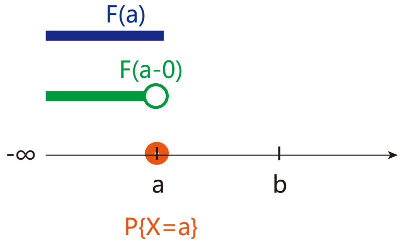
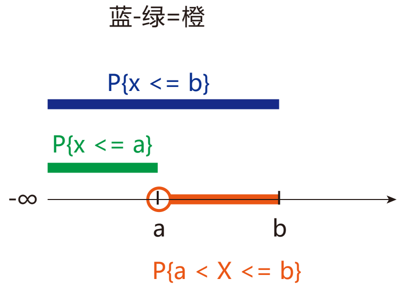
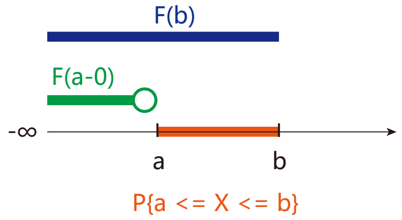




$\lim_{h\to 0}\frac{F(x+h)-F(x)}{h}$ = 累积函数F,在x点处的斜率（即导数）

或者说，它反映了概率在 x 点处"密集程度".

3 ★累加函数的计算公式

累加函数 $F(x) = PX \leq x$ 的公式有	图中：蓝-绿=橙
(1) $P\{X \leq a\} = F(a)$	
(2) $P\{X < a\} = F(a-0) \leftarrow$ 其中的 F(a-0)：就是从左边逼近a，不包括a点．所以是“左极限”．就是(-∞,a)这段区间的概率之和，不包括a点上的概率．	
(3) $P\{X > a\} = 1 - P\{X \leq a\} = 1 - F(a)$	
(4) $P\{X \geq a\} = 1 - F(a-0)$	
(5) $P\{X = a\} = F(a) - F(a-0)$	
(6) $P\{a < X \leq b\} = P(X \leq b) - P(X \leq a)$	
(7) $P\{a \leq X \leq b\} = F(b) - F(a-0)$	

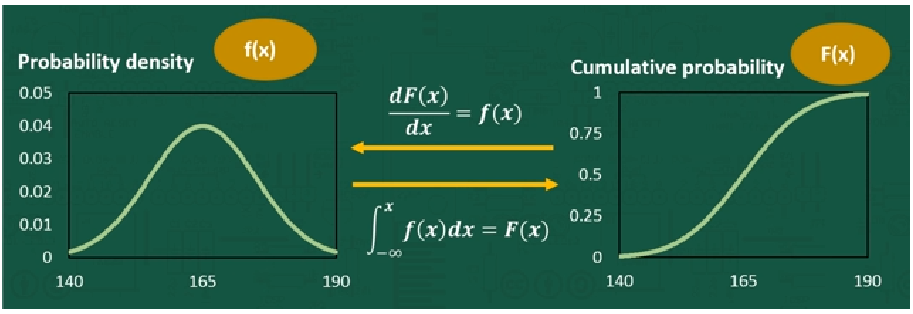
---

4 对“概率函数f(x)”求积分，就得到“累加函数F(x)”．反之，对“累加函数F(x)”求导，就得到“概率函数f(x)”．即：① $\int$ 概率函数 $f(x) =$ 累加函数 $F(x)$，② $(\text{累加函数} F(x))' = \text{概率函数} f(x)$

对“概率函数f(x)”求积分，就得到“累加函数 F(x)”

对“累加函数 F(x)” 求导，就得到“概率函数f(x)”．

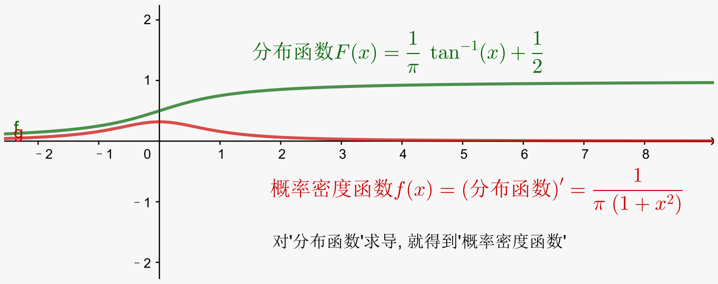
4 对“概率函数$F(X)$”求积分, 就得到“累加函数$F(X)$”. 反之, 对“累加函数$F(X)$”求导, 就得到“概率函数



例

$$\underbrace{f\left(x\right)}_{\text{概率函数(相当于导函数)}} = \frac{1}{\pi\left(1+x^2\right)}$$

$$\underbrace{F\left(x\right)}_{\text{累加函数}} = \int_{-\infty}^{x} \underbrace{f\left(t\right)}_{\text{概率函数}} dt = \int_{-\infty}^{x} \underbrace{\frac{1}{\pi\left(1+t^2\right)}}_{\text{导函数}} dt = \frac{1}{\pi}\left(\arctan|_{-\infty}^{x}\right) = \frac{1}{\pi}\arctan x + \frac{1}{2}$$



例

有累加函数 $F\left(x\right) = \begin{cases} 0 & (x<0) \\ Ax^2 & (0\le x<1) \\ 1 & (x\ge 1) \end{cases}$

$\rightarrow$ 求常数A : $\underbrace{\lim_{x\to 1^-} Ax^2}_{\text{从}x=1\text{的左侧,逼近}x=1} = A\left(1\right)^2 = A = \underbrace{F\left(1\right)=1}_{\text{因为当}x\ge 1\text{时},F(x)=1}$

$\rightarrow$ 求概率函数f(x) : $f\left(x\right) = F'\left(x\right) = \begin{cases} 0'=0 & (x<0) \\ \left(Ax^2\right)' = \left(x^2\right)' = 2x & (0\le x<1) \\ 1'=0 & (x\ge 1) \end{cases}$

$\rightarrow$ 求 :

$$P\left\{0.3<X<0.7\right\} = \underbrace{F\left(0.7\right)}_{=\int_{-\infty}^{0.7} f(t)dt} - \underbrace{F\left(0.3\right)}_{=\int_{-\infty}^{0.3} f(t)dt}$$

因为在$0\le x<1$的区间上,$F\left(x\right) = Ax^2$,而其中的A我们上面已经算出 $=1$,

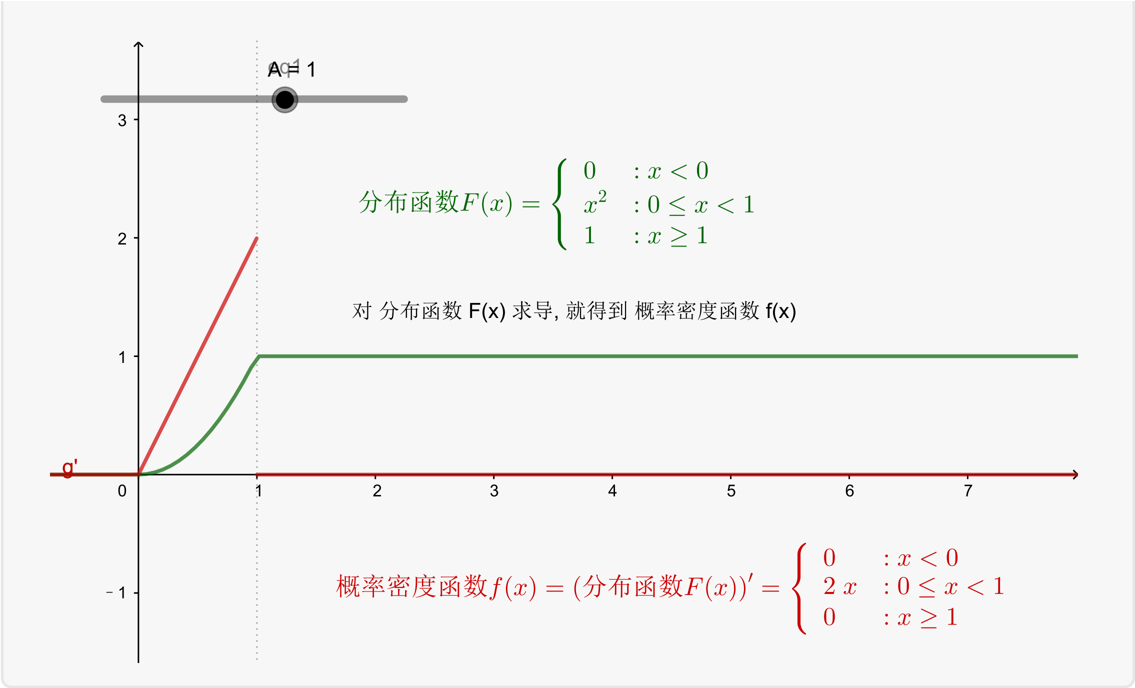
所以$F\left(x\right) = Ax^2 = \left(1\right)x^2 = x^2$

所以: $F\left(0.7\right) = \left(0.7\right)^2 = 0.49$

$F\left(0.3\right) = \left(0.3\right)^2 = 0.09$

因此: $P\left\{0.3<X<0.7\right\} = F\left(0.7\right) - F\left(0.3\right) = 0.49-0.09 = 0.4$

事实上,本例的 $P\left\{0.3<X<0.7\right\} = \int_{0.3}^{0.7} \underbrace{\left(2x\right)}_{\text{即概率函数}f(x)} dx$

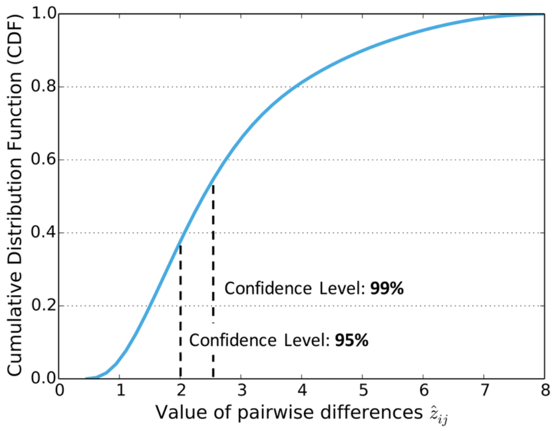


---

5 性质

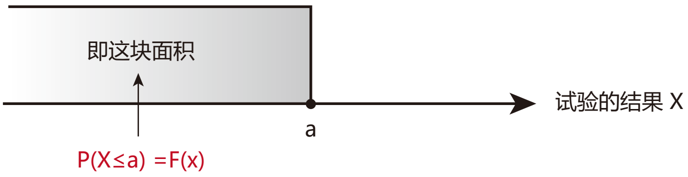
5.1 性质1: 有界性. $F(x) = P\{随机变量X \le 随机变量的取值x\}, \ x \in (-\infty, +\infty)$. 即 0 $F(x) \le 1$

累加函数(CDF) F(x), 就是一个普通的实函数. 其定义域是 $x \in (-\infty, +\infty)$. 值域是 $y \in [0, 1]$.

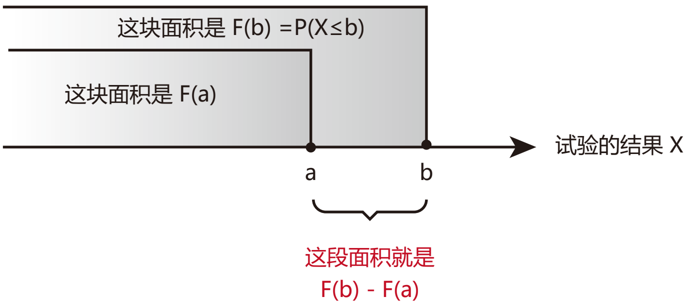


累加函数：

$$\boxed{F(x) = P(X \le a) = \int_{lower\lim it}^{a} f(x)\,dx}$$



$$\boxed{P(a \le X \le b) = P(a < X < b) = \int_a^b f(x)\,dx = F(b) - F(a)}$$



$$P(x_1 < X \le x_2) \text{ ¸ 对于随机变量}X\text{在}(x_1, x_2]\text{这段区间上的概率,它的值}$$
$$= F(x_2) - F(x_1)$$
$$= P\{X \le x_2\} - P\{X \le x_1\}$$

对于"连续型随机变量", 有没有两端的端点, 无所谓, 不影响概率值(因为它在任何一个"确定点"的概率都是0嘛). 即:

$$P\{a \le X \le b\}$$
$$= P\{a < X \le b\} \quad \leftarrow \text{即, 两端是否有"等于号", 无所谓.}$$
$$= P\{a \le X < b\}$$
$$= P\{a < X < b\}$$

同样:
$P\{X < a\} = P\{X \le a\} \leftarrow$ 有没有"等于号"无所谓
$P\{X > a\} = P\{X \ge a\}$

5.2 性质2: 单调不减性. 即对于任意的 $x_1 < x_2$, 有: $F(x_1) \le F(x_2)$

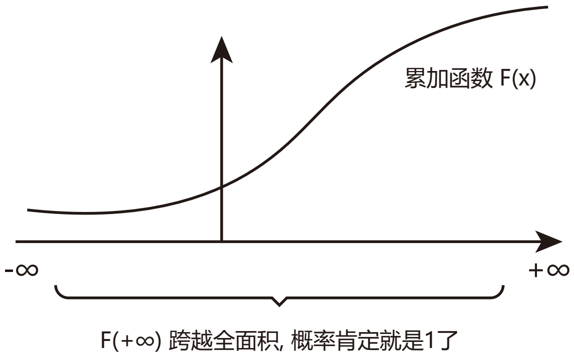
F(x)是关于x的"不减函数", 类似于"单调递增"的概念. "不减"的意思就是, 该函数的y值不会下降, 只会"增长"或"平移向前".

比如, "分数小于等于70分的人" 其概率一定是小于等于 "分数小于80分的人". 即 $F(70) \le F(80)$.

5.3 性质3: 规范性. F(-∞)=0 , F(+∞)=1

$\boxed{\underbrace{F(-\infty)}_{=P(X\le-\infty)} = \lim\limits_{x\to-\infty} F(x) = P(X < -\infty) = P(\Phi) = 0}$ ← 称之为 "不可能事件".
如果随机变量X的取值, 比 -∞还小, 那其概率, 就只能是0了.

$\boxed{\underbrace{F(+\infty)}_{=P(X\le+\infty)} = \lim\limits_{x\to+\infty} F(x) = P(X < +\infty) = P(\Omega) = 1}$ ← 称之为 "必然事件".
如果随机变量X的取值, 在 +∞以下, 那其概率, 肯定就是100%了, 就是1.



5.4 性质4：右连续性. $\lim_{x\to x_0^+} F(x) = F(x_0)$

这个等式的意思就是说: 累加函数在x_0点的右极限, 就等于累加函数在该点处的函数值.

右连续	所谓"右连续", 就是"函数从x在某点的右侧, 逼近该点"的极限值, 就等于"该点处的y值", 即: $\lim_{x\to a^+} F(x) = F(a)$.
左连续	同理, "左连续"就是: $\lim_{x\to a^-} F(x) = F(a)$.
连续	同时满足"左连续"和"右连续"的函数, 就称为是"连续"的. 即 $\lim_{x\to a} F(x) = F(a)$.

满足上面4条性质的, 就一定是"累加函数". 反之, "累加函数"也一定有这4条性质.

第四部分　随机变量函数的分布

意思就是说, 假如我们已经知道某个X 是某种类型的分布了, 比如 X 它是几何分布的, 二项分布的等. 则进一步, 而我们还想知道, 用这个X 来构造出的其他函数, 会是什么类型的分布呢? 比如, Y=3X-5, 这个Y是由X构造出来的, 那么这个Y, 也是和X相同类型的分布吗? 还是说, Y是其他类型的分布?

→ 随机变量X, 它取x时, 其"累加函数"是: $\boxed{F_X(x) = P\{X \le x\}}$

→ 由随机变量X, 构造出的一个新 Y (比如 Y="多少倍的X, 再加上某个数"之类), 这个Y 的"累加函数", 是: $\boxed{F_Y(x) = P\{Y \le x\}}$ ← 等号左边的 $F_Y(x)$ 意思是: Y是从X构造出来的. 累加函数(用F表示), 所以 $F_Y(x)$ 就是指 "由X构造出来的新的Y"的累加函数.